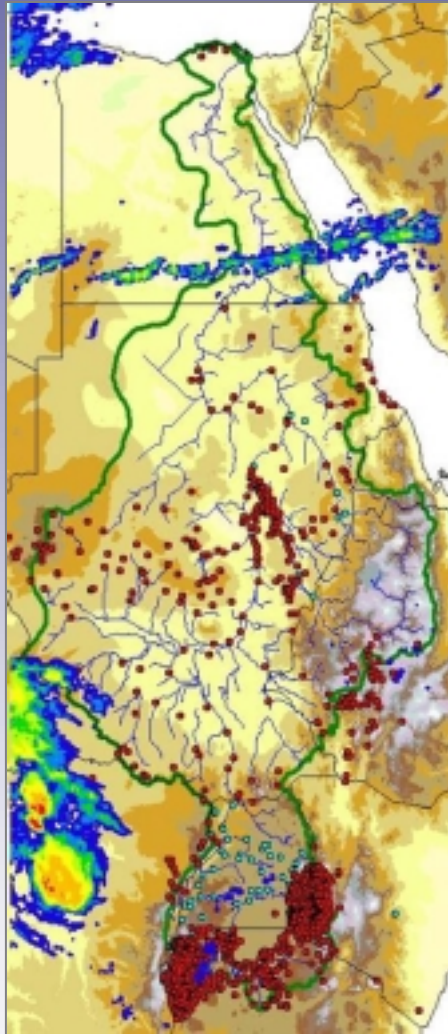




Nile Decision Support Tool Watershed Hydrology

Burundi
Congo
Egypt
Eritrea
Ethiopia
Kenya
Rwanda
Sudan
Tanzania
Uganda



Developed collaboratively by

The Nile Basin Nations,

**The Georgia Water Resources Institute
at the Georgia Institute of Technology,**

and

**The Food and Agriculture Organization
of the United Nations**

June 2003



Nile Decision Support Tool (Nile DST) Watershed Hydrology

Report developed by

Amy Tidwell

Research Associate

Lori Visone

Research Associate

Aris Georgakakos

Project Director

Georgia Water Resources Institute
School of Civil and Environmental Engineering
Georgia Institute of Technology

In collaboration with

The Nile Basin Nations

and

The Food and Agriculture Organization (FAO)

of the United Nations

Nile Basin Water Resources Project (GCP/INT/752/ITA)

June 2003

Acknowledgements

This report and associated software were developed by the Georgia Water Resources Institute (GWRI) at the Georgia Institute of Technology as part of the Nile Basin Water Resources Project (GCP/INT/752/ITA). This project was funded by the Government of Italy and was executed for the Nile Basin nations by the Food and Agriculture Organization (FAO) of the United Nations.

The GWRI Director and project staff are grateful to the Nile Basin nations (Burundi, Congo, Egypt, Eritrea, Ethiopia, Kenya, Rwanda, Sudan, Tanzania, and Uganda), their focal point institutions, their Project Steering Committee (PSC) members, and their National Modelers for entrusting us to work with them in this important basin-wide project. The development of databases, models, technical reports, software, and user manuals are key but not the only project accomplishments. Even more important are the evolving contributions relating to people and the difference the project is poised to make in data and information sharing, developing a common knowledge base for policy debates, and long term capacity building.

GWRI is also grateful to the Government of Italy and to FAO for sponsoring this project and for providing dependable logistical and technical support through the FAO offices in Rome and Entebbe.

It is our hope that the Nile DST effort will contribute in some positive way to the historic process of the Nile Basin nations to create a sustainable and peaceful future.

Aris Georgakakos
GWRI Director
Atlanta, June 2003

Disclaimer and Copyright Notice

The contents of this report do not necessarily reflect the views of the Nile Basin nations or those of the Government of Italy and FAO.

The Nile Basin Nations shall have ownership of the deliverable application software and of the information generated under this contract. However, GWRI and Georgia Tech shall reserve all intellectual property rights of the methods and decision support technology utilized and developed under this contract. In keeping with standard professional practices, publications containing results of the Nile DST software, reports, and manuals shall acknowledge the original information source and reference its authors.

TABLE OF CONTENTS

NILE WATERSHED HYDROLOGIC MODELING	1
1. OVERVIEW	1
2. HYDROLOGIC MODEL.....	2
2.1 Methodology.....	2
2.2 Model Training.....	3
2.3 Model Verification.....	4
3. WATERSHED DELINEATION	6
4. NZOIA WATERSHED	11
4.1 Precipitation Estimation	11
Station Selection.....	12
MAP Estimation.....	12
Zonal MAP Estimation.....	13
Results	15
4.2 Temperature Estimation.....	20
4.3 Streamflow Record.....	20
Station Selection.....	20
Rating Curves	21
Station 60079.....	21
Station 60082.....	22
Station 60077.....	23
Streamflow Data Analysis and Extension.....	25
4.4 Model Calibration and Results	27
5. KAGERA WATERSHED	30
5.1 Precipitation Estimation	30
Station Selection.....	30
MAP Estimation.....	32
Normal Weight MAP Estimation.....	32
Results	35
5.2 Temperature Estimation.....	38
Station Selection.....	38
Temperature Series Development.....	40
Results	43
5.3 Streamflow Record.....	45
Station Selection.....	45
Streamflow Data Analysis and Extension.....	45
Data Extension	45
5.4 Model Calibration and Results	46
5. OUTSTANDING DATA ISSUES.....	49

Nile Watershed Hydrologic Modeling

1. Overview

The analyses and results presented in this report were carried out as part of the development of the Nile Decision Support Tool (Nile DST) watershed hydrology component. The purpose of the hydrologic models is to describe the response of the Nile watersheds (streamflow and soil moisture) to various conditions of atmospheric forcing (rainfall and temperature). Methods for watershed delineation, data analysis and preparation, and hydrologic modeling are presented and discussed.

The methodology for the hydrologic model used in the Nile DST is described in Chapter 1. Subsequent chapters detail the tools and procedures applied to specific sub basins as part of the watershed assessments. Before hydrologic analysis could begin it was necessary to subdivide the Nile Basin into appropriate hydrologic units. Chapter 2 describes the technical tools and procedures used for watershed delineation.

Chapters 3 and 4 detail the extensive data analysis and processing applied to specific sub basins. Topics presented in these chapters include: assessment of existing ground station data, mean areal precipitation estimation, temperature estimation, analysis of streamflow records, and calibration and verification of the hydrologic model. Finally, Chapter 4 describes outstanding data issues identified during data analysis and hydrologic modeling. It is clear that meaningful hydrologic modeling requires adequate meteorologic and hydrologic data that have not been assembled for several Nile watersheds. As part of a follow-up project phase, it is **recommended** that such (existing) data be assembled, quality controlled, and used to (re)calibrate adequate hydrologic models. To this end, the procedures presented herein can be used to guide this follow-up process, step by step.

The hydrologic data, data preparation tools, and models are part of and can be accessed through the Nile DST.

2. Hydrologic Model

Analysis and modeling for the Nile Basin watersheds utilizes a hydrologic model that was developed at the Georgia Institute of Technology (Georgia Tech). Model methodology and development is described here, including the training and verification procedures.

2.1 Methodology

The basic model principle is water balance. Let k denote a particular time increment (month or 10-day period), $s(k)$ the watershed soil moisture storage at the beginning of period k , $P(k)$ the precipitation during interval k , $PET(k)$ the potential evapotranspiration, and $Q(k)$ the streamflow. Then, the watershed water balance can be stated as follows:

$$s(k + 1) = s(k) + P(k) - PET(k)[\alpha s(k) + \beta] - \gamma s(k) - Q(k),$$

where α , β , and γ are model parameters for calibration. In addition to streamflow, this formulation includes two additional loss terms. The first represents evapotranspiration occurring at the potential rate and applied to the current moisture storage, and the second represents percolation to groundwater aquifers as a linear function of soil moisture. Potential evapotranspiration (PET) is estimated based on the temperature over the model time increment (monthly or 10-day) using the following equation (Dingman, 1994):

$$PET = 0.409 \times 6.11 \times \exp\left[\frac{17.3 \times T}{T + 237.3}\right],$$

where T is mean monthly temperature in degrees C, and PET is in centimeters per month.

The water balance equation requires a relationship between streamflow and the known quantities at time k . The model utilizes the water balance equation to develop a soil moisture index, rather than soil moisture itself, that can be related to streamflow. For an ideal streamflow index IQ , $IQ_1 \leq IQ_2$ implies and is implied by $Q_1 \leq Q_2$ for all $[IQ_1, IQ_2]$ and $[Q_1, Q_2]$ pairs in the ranges of IQ and Q (Georgakakos and Yao, 2000). Further, this implies that IQ_2 and Q_2 occur with the same frequency so that

$$\text{Pr obability}\{IQ \leq IQ_2\} = \text{Pr obability}\{Q \leq Q_2\}.$$

Motivated by this concept, the model uses frequency matching to estimate model streamflow. The streamflow index used by the model is defined as follows:

$$IQ(k) = \delta s(k) + (1 - \delta) P(k),$$

where δ is a model parameter defined on the interval $[0,1]$. Namely, the streamflow index is considered to be a function of the rainfall and the current soil moisture storage. During model

training, frequency curves are generated for both the index IQ and streamflow Q. At each simulation time step, the model determines the frequency of the index variable, IQ(k), and then retrieves the historical streamflow, Q(k), that occurs with the same frequency. This is shown graphically in the figure below.

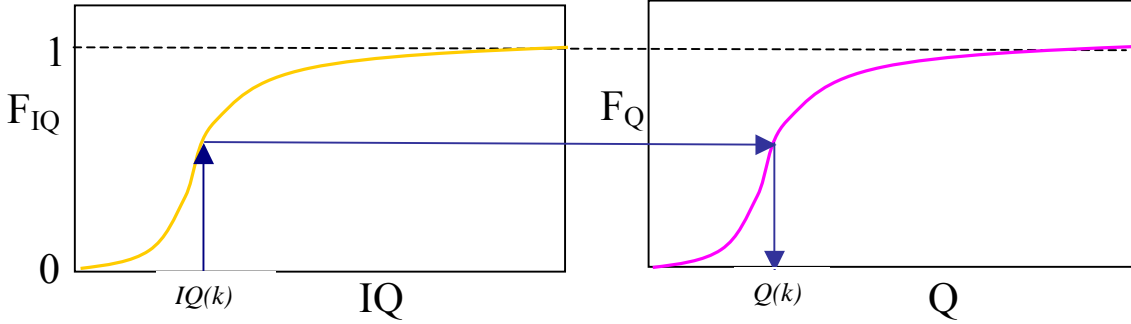


Figure 2.1: Frequency Matching

Thus, the model includes four parameters α , β , γ , and δ that must be calibrated for each watershed. The parameter calibration process uses the split sample approach in which the period of record containing data for precipitation, temperature, and streamflow is split into two sections. The first section is used for model training, while the second is used to verify model results. The calibration process is discussed below.

2.2 Model Training

The data sequences (P, PET, and Q) in the first data sample are used as input for the recursive application of the water balance equation. The procedure is initialized by selecting an initial model parameter set $[\alpha_0, \beta_0, \gamma_0, \delta_0]$ and an initial soil moisture, $s(0)$. The initial soil moisture can be set equal to an appropriate average soil moisture value.

$$s(1) = s(0) + P(0) - PET(0) [\alpha_0 s(0) + \beta_0] - \gamma_0 s(0) - Q(0)$$

$$s(2) = s(1) + P(1) - PET(1) [\alpha_0 s(1) + \beta_0] - \gamma_0 s(1) - Q(1)$$

•
•
•

$$s(N) = s(N - 1) + P(N - 1) - PET(N - 1) [\alpha_0 s(N - 1) + \beta_0] - \gamma_0 s(N - 1) - Q(N - 1)$$

Once the soil moisture sequence has been generated, the streamflow index can be computed as follows:

$$\begin{aligned}
 IQ(1) &= \delta_0 s(1) + (1 - \delta_0) P(1) \\
 IQ(2) &= \delta_0 s(2) + (1 - \delta_0) P(2) \\
 &\vdots \\
 &\vdots \\
 &\vdots \\
 IQ(N) &= \delta_0 s(N) + (1 - \delta_0) P(N)
 \end{aligned}$$

The IQ and Q sequences are ranked and their respective frequency curves developed. A frequency plotting formula (e.g., the Cunane formula) can be used to calculate frequency position based on N observations for a variable X:

$$F_x = \text{Probability}\{X \leq x\} = 1 - \frac{m_x - 0.4}{N + 0.2},$$

where m_x is the rank of observation x among the N available observations ranked in descending order.

2.3 Model Verification

Once the IQ and Q frequency curves have been developed during the training procedure, probability matching can be employed, as described previously, to generate simulated flows for the verification sample and any other future sequences. The soil moisture from the last step of the training period is used as the initial soil moisture input for the verification period. This soil moisture is used to initiate the simulation process as follows:

1. $s(0) = s(N)$;
2. $IQ(0) = \delta_0 s(0) + (1 - \delta_0) P(0)$;
3. Using the frequency correspondence, $F_{Q(0)} = F_{IQ(0)}$, select $Q(0)$ from the training sample;
4. Determine $s(1) = s(0) + P(0) - PET(0) [\alpha_0 s(0) + \beta_0] - \gamma_0 s(0) - Q(0)$.

Subsequent flows are generated recursively by computing $IQ(k)$, finding the frequency $F_{IQ(k)}$ with respect to the training frequency distribution, selecting $Q(k)$ from the streamflow frequency curve that corresponds to $F_{IQ(k)}$, and finally repeating the procedure for the next time step.

Once the simulated sequence has been produced, model generated flows are compared to observed flows. The model error is evaluated as the square difference of the observed and simulated streamflow:

$$\text{Total Error} = \sum_k (Q_{(k)} - Q_{(\text{observed})})^2$$

where k spans the data of the second (verification) sample.

The training and verification calculations are repeated, each time specifying a new set of model parameters $[\alpha_i, \beta_i, \gamma_i, \delta_i]$, until the model error is minimized in the verification sequence, yielding the optimal parameter set $[\alpha^*, \beta^*, \gamma^*, \delta^*]$.

Model calibration and verification applications are provided at later sections of this report.

3. Watershed Delineation

In this section, we discuss the calibration of the hydrologic model for the Nzoia and Kagera Basins. The process is described in detail to illustrate the data issues and the need to carry it out in collaboration with country engineers and hydrologists.

Delineation of the Nile Basin watersheds is necessary to create the hydrologic units for which the hydrologic model is to be developed. Watersheds are delineated for the Lake Victoria region, Sobat, Blue Nile and Atbara rivers. Several iterations were performed before hydrologically consistent watersheds could be developed. This process is described below.

The initial input data was obtained as Arc Info coverages and converted to Arc View shape files. The data used for watershed delineation includes the Nile Basin digital elevation model (DEM) [Niledem_dd] and the rivers network [rivers_dd]. The DEM grid size is 30 arc seconds. The initial watersheds were developed by “burning” the river network into the DEM using the watershed delineator extension (WshdDel.avx) of Arc View. The watersheds produced were not appropriate because of the inconsistent resolutions of the DEM and of the river network. For example, during the burning process in several locations two different river segments ended in the same DEM grid. This resulted in the rivers being joined in the headwaters and thus a river segment that flows uphill (Figure 3.1). Also, watershed outlets did not coincide with lake boundaries, but continued into the lakes (Figure 3.2).

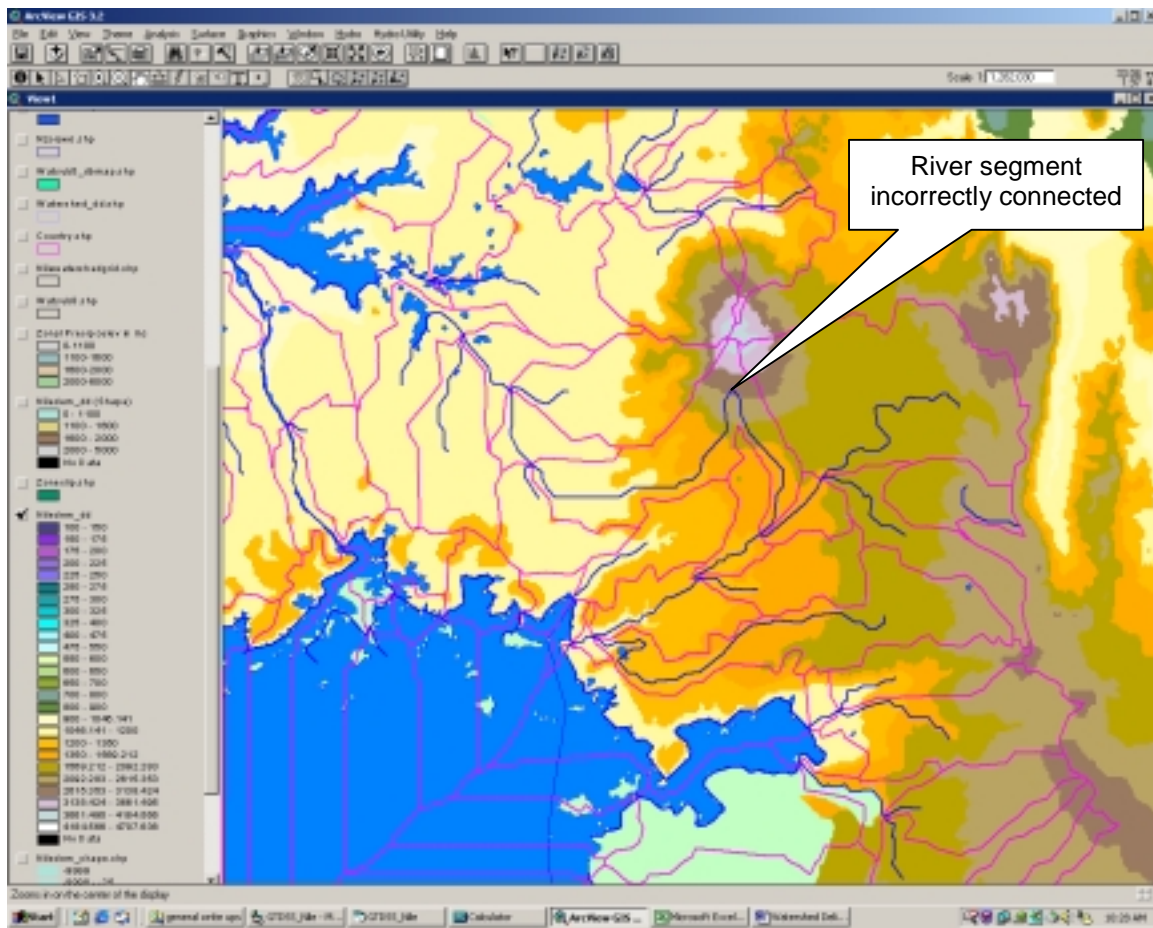


Figure 3.1: Example of River Segment Connection

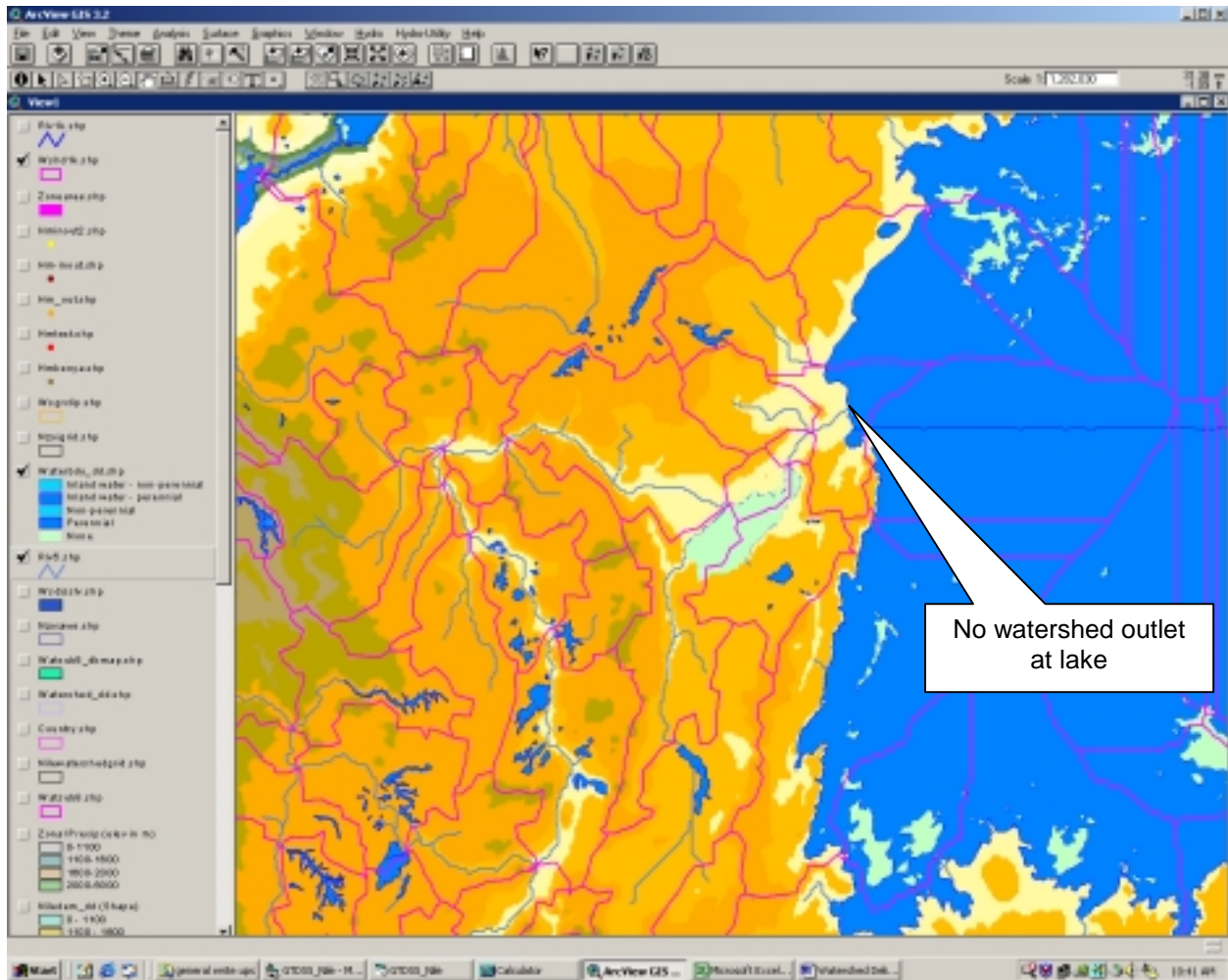


Figure 3.2: Example of Watershed Outlets at Lake Boundaries

To correct the resolution discrepancies, the river network shape file was reviewed in detail and trimmed in areas where adjacent segments could result in an incorrect merging of headwater streams. Floating streams, streams not connected to a downstream segment, were also deleted or connected. The connection was made if visual analysis of the DEM indicated a likely connection to another river within a short distance. To address the watershed outlet problem, the watershed delineation method was switched from the watershed delineator extension to the watershed delineation method in SWAT (Soil and Water Assessment Tool). SWAT is a water quality modeling extension for Arc View that has better and more functional watershed delineation tools. Figures 3.3 and 3.4 show the results of these corrections.

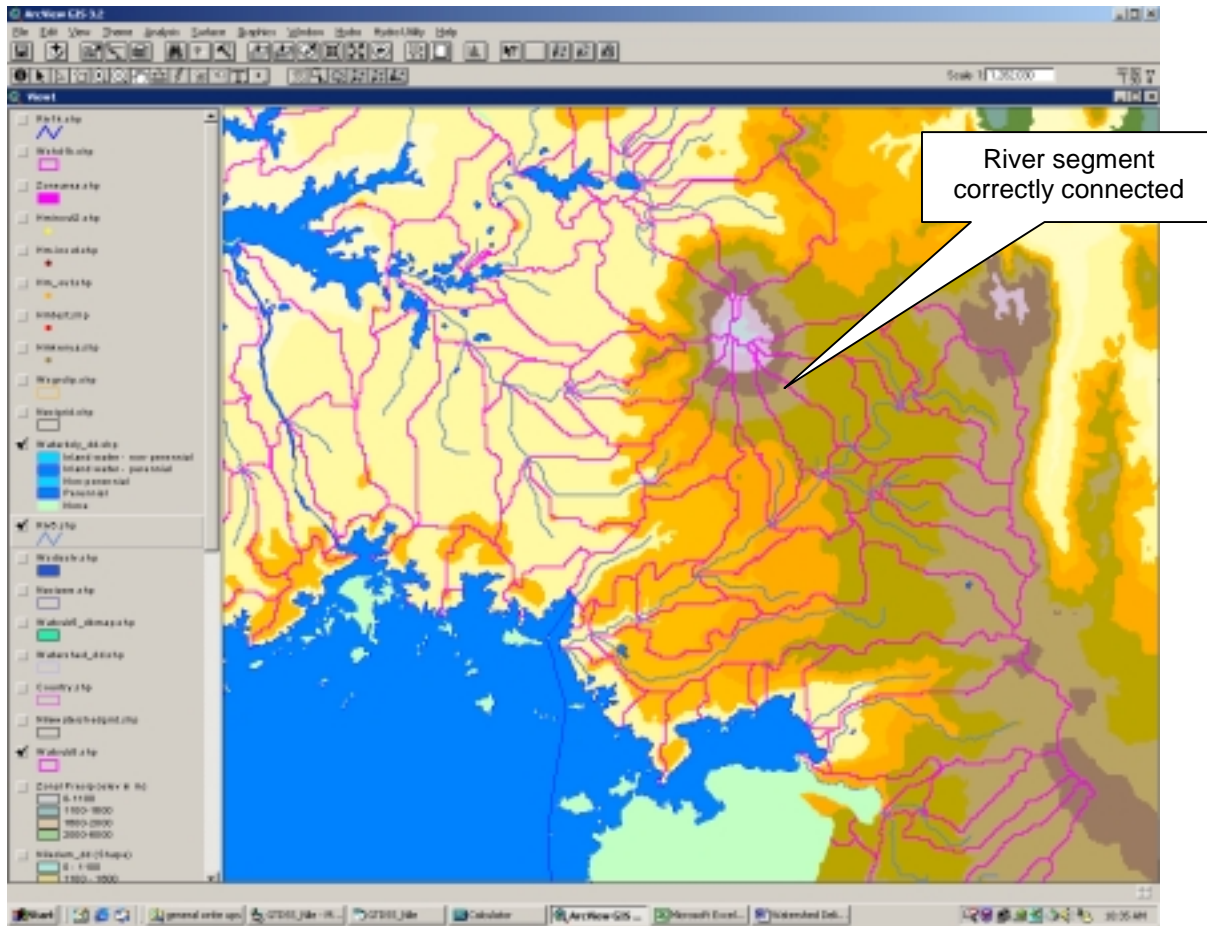


Figure 3.3: Example of Corrected River Segment Connection

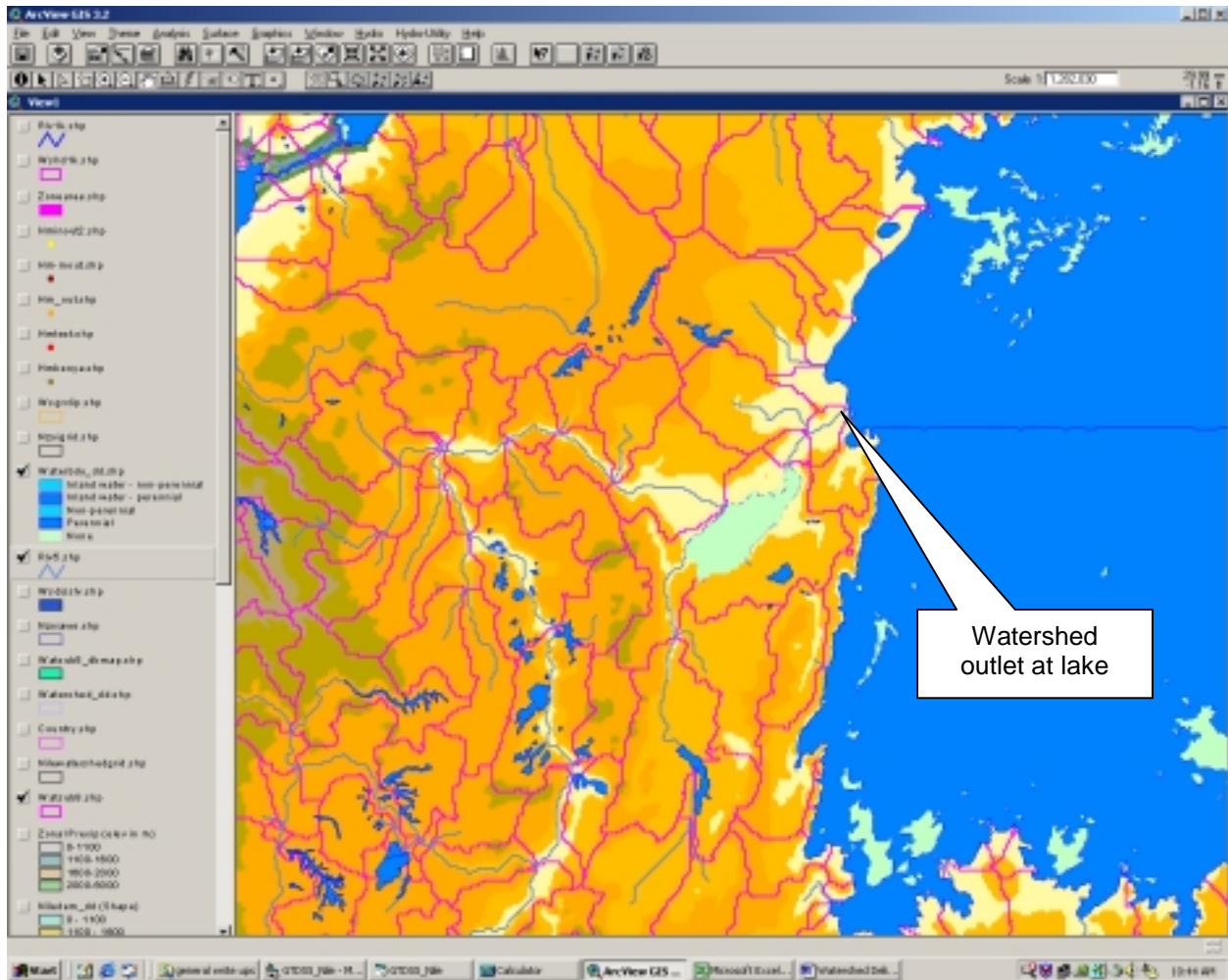


Figure 3.4: Example of Watershed Outlet Correction at Lake Boundary

The watershed delineation process is based on a grid cell threshold, which is selected by the user; however, an absolute minimum threshold is established by the delineation software. When the delineation algorithm determines that the number of grid cells (DEM grids) flowing to a point has exceeded the threshold, a watershed is delineated. The initial delineation using the SWAT software was based on 1000 grid cells. Also, a point file of outlets into major lakes was developed. This point file indicates the watershed outlets to the delineation software and resolves the second problem of watersheds not coinciding with lake boundaries. These watersheds were reviewed in detail and additional revisions were made to the river network to realistically represent the watersheds.

The final delineation was performed using the revised river network and by selecting the smallest threshold (i.e., 559) for the grid cells. The watershed GIS layer developed from this delineation process was added to the Nile DST interface. The average watershed size in this layer is approximately 800 square kilometers. Larger watersheds can be delineated at user-selected river nodes by con-joining all upstream sub-watersheds.

4. Nzoia Watershed

The Nzoia Watershed is located in Kenya and drains into Lake Victoria (Figure 4.1). The Nzoia Basin is approximately 15,000 km².

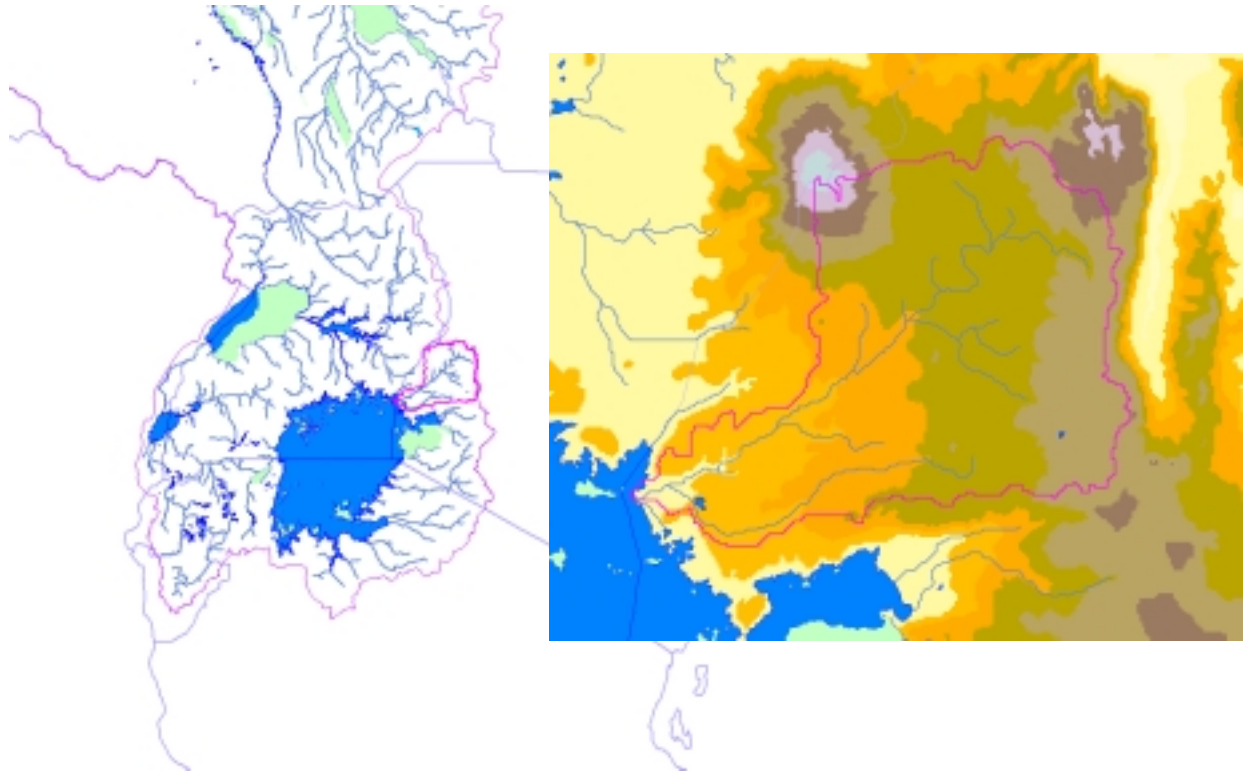


Figure 4.1: Nzoia Watershed

4.1 Precipitation Estimation

The precipitation input files for the watershed hydrology model were developed using two methods. The first method estimates the mean areal precipitation (MAP) of the whole watershed. The second estimates the mean areal precipitation for individual watershed zones corresponding to different mean ground elevations and combines them according to their relative areas (zonal MAP). The precipitation stations used in the model were selected using analysis features available in the Nile DST. The time frame selected was based on the flow records available for the watershed to be modeled. Both 10-day and monthly precipitation input files were developed.

Station Selection

The station selection process was the same for both MAP procedures. The previously delineated watersheds were turned on in the map view of the Nile DST. The sub watersheds that make up the watershed of interest were selected. The ground-monitored stations with precipitation data contained in the watershed were selected using the 'selected by theme' feature. This feature highlights the stations with precipitation data in the data tree. From the data tree these stations were charted (2.2). The data was charted using the yearly statistic of data count and time interval. Stations were selected if records were 90% complete (~330 days) for the majority of the time period of interest.

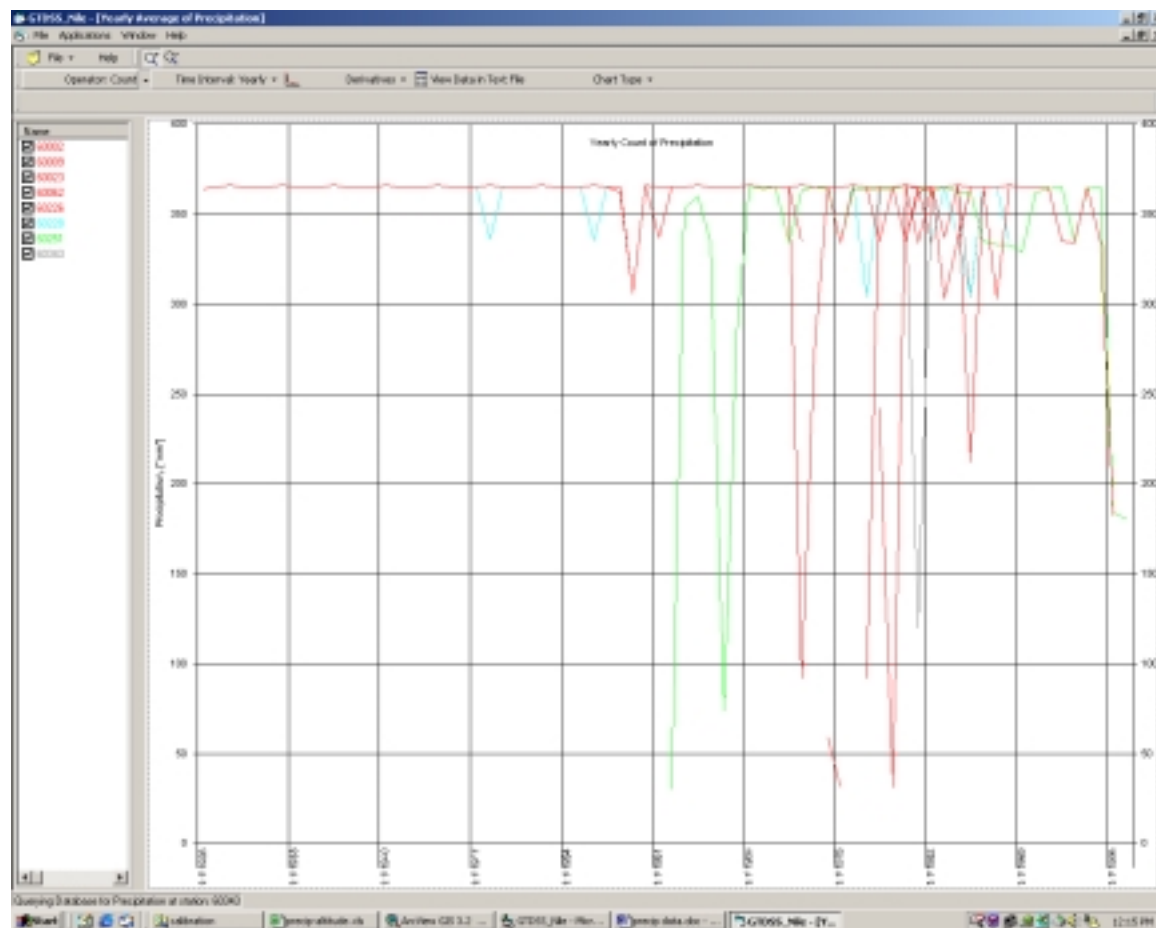


Figure 4.2: Precipitation Data in Yearly Count Chart

MAP Estimation

For the first MAP calculation, the Data Analysis option in the Database Application was used. An analysis tree was developed to perform an inverse distance weighting (IDW) using a grid, watershed and precipitation stations as input (2.3). The selected grid is available in the Nile DST and was developed for use in watershed modeling. The grid has a cell size of 0.1 degree. The

previously selected watershed is used for the watershed input, and the stations selected in the last section are used as the precipitation input.

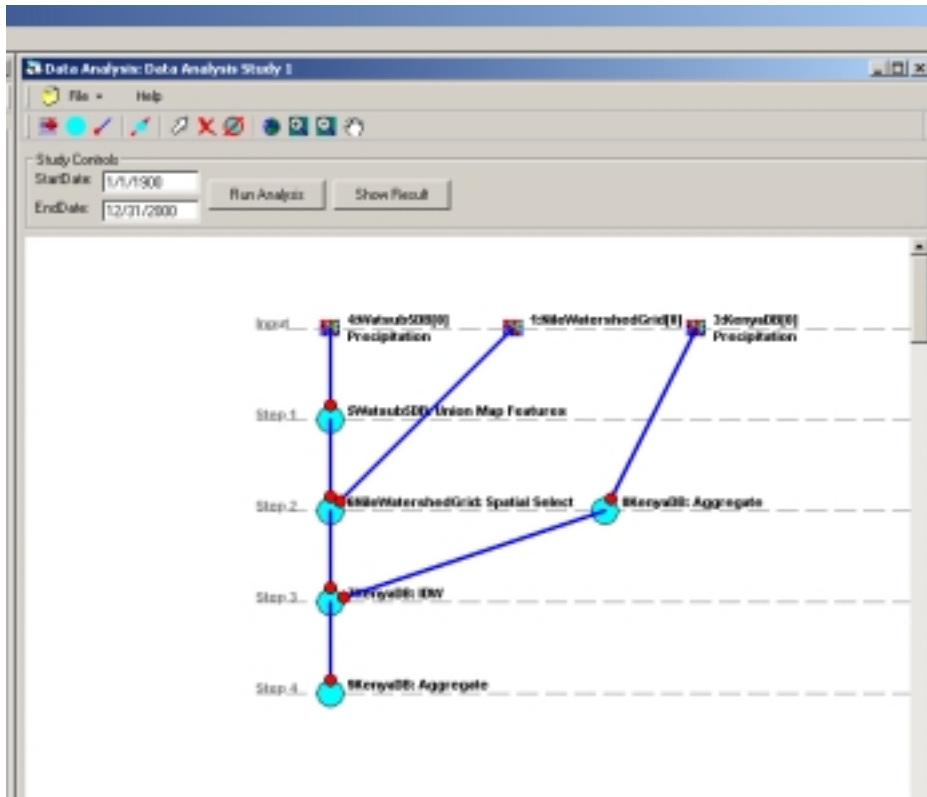


Figure 2.3: Data Analysis Tree for MAP Calculation

The time interval for the analysis is chosen and the ‘run analysis’ button is selected. The IDW calculation is performed for each grid cell in the watershed. As a result, each grid has a time sequence of precipitation data. This calculation includes a weighting based on the grid cell size since some grids may not be completely contained within the watershed. Next the grid data is aggregated spatially, so that a single daily time series is developed for the entire watershed. The last node in the tree may be changed so that the temporal aggregation is either 10-day or monthly.

Zonal MAP Estimation

Many features, both climatic and geographical, affect the distribution of rainfall. Due to orographic and lake effects, a second set of precipitation data was developed using a zonal methodology. The development of zonal MAP required additional analysis of the precipitation data prior to construction of the data analysis tree, in order to define the zones. The yearly average precipitation for each station was exported as a text file from the Nile DST chart. The average over the entire record was graphed as a function of elevation (Figure 4.4).

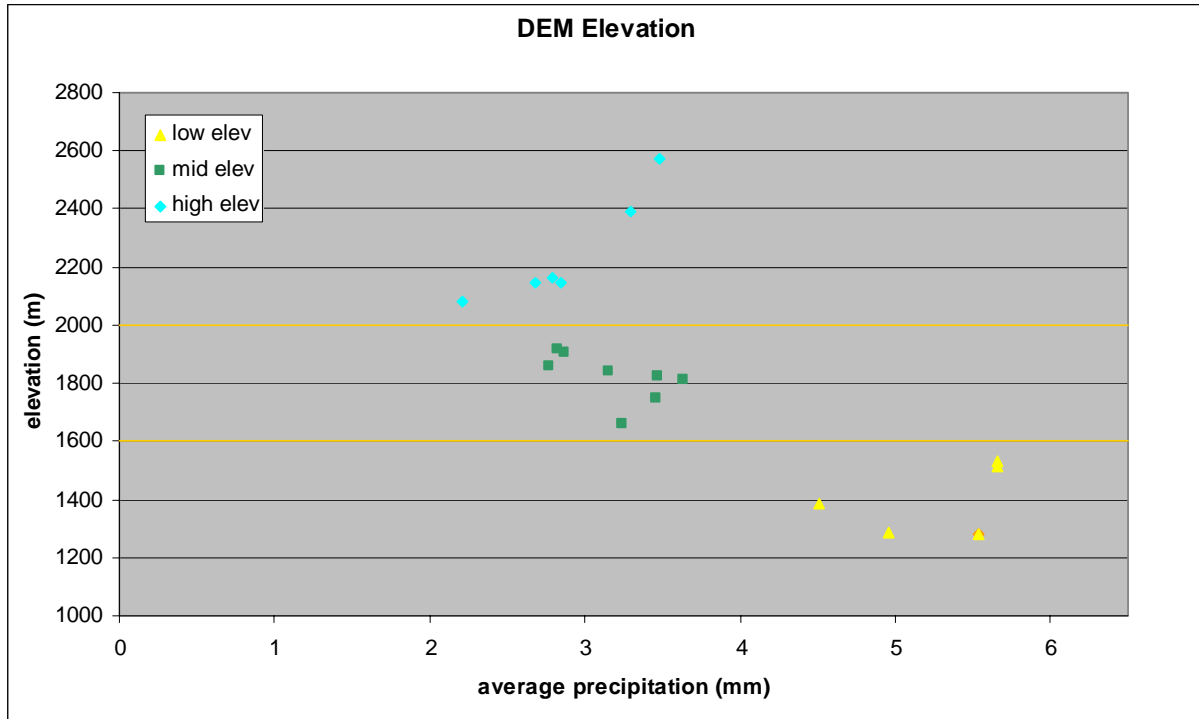


Figure 4.4: Development of Precipitation Zones based on Elevation

This basin shows three distinct zones of precipitation. As a result a zonal map was developed in Arc View of these three zones (Figure 4.5).

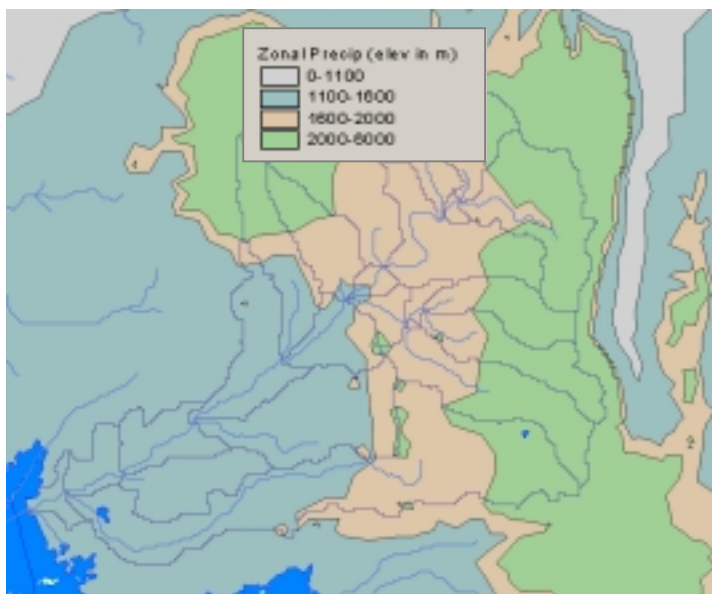


Figure 4.5: Zonal Precipitation based on Elevation

This map was imported into the Nile DST for use in the zonal analysis. The same data tree developed previously for the MAP calculation may be modified for use in the zonal MAP calculation. The only modification required was to change the IDW node to a zonal IDW node. To complete this change a shape file must be selected that delineates the zones to be used for the

zonal MAP calculation. The MAP calculation is the same as before except a time series of precipitation for each grid cell is developed using the IDW method with only the stations in the same zone as the grid cell. Once again, these results may be aggregated for both 10-day and monthly time intervals by changing the last node in the data tree accordingly.

Results

For the Nzoia watershed, the MAP was produced using 19 stations that contained data corresponding to the time interval for which flow information was available (roughly 1963 to 1990). Figure 4.6- shows the Nzoia basin and the precipitation stations. Not all stations had records for the entire interval, but the stations selected generally had 90% of daily records per year for the station recording period.



Figure 4.6: Nzoia Watershed and Precipitation Stations

To calculate the zonal MAP, a zonal grid was developed for the watershed (Figure 4.7). The Nzoia precipitation records showed three distinct zones based on altitude; zone 3 (less than 1600m), zone 2 (1600-2000m) and zone 1 (greater than 2000m).

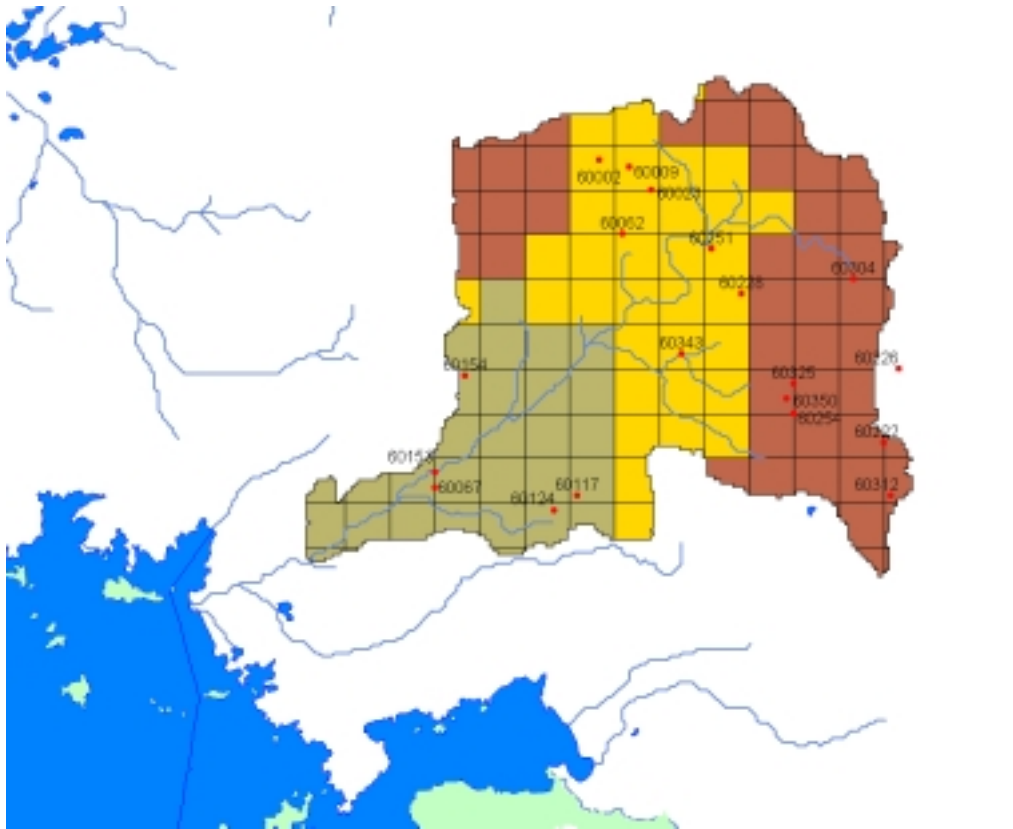


Figure 4.7: Zonal Grid for Nzoia Watershed

Both the MAP and Zonal MAP were calculated using the methods previously discussed. The resulting 10-day and monthly values for MAP and Zonal Map are plotted in Figures 4.8 and 4.9. The difference between the MAP and the Zonal MAP for both the 10-day and the monthly is plotted in Figures 4.10 and 4.11. Figures 4.12 and 4.13 show the monthly and 10-day precipitation for each zone.

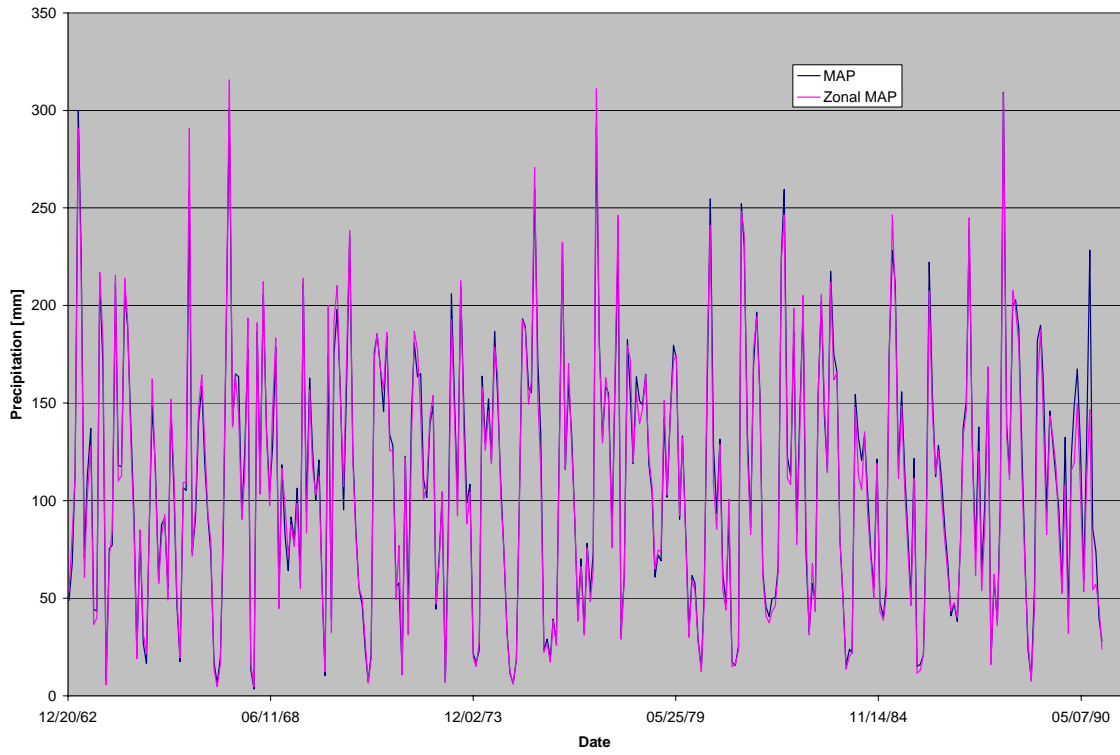


Figure 4.8: Monthly MAP and Zonal MAP for Nzoia

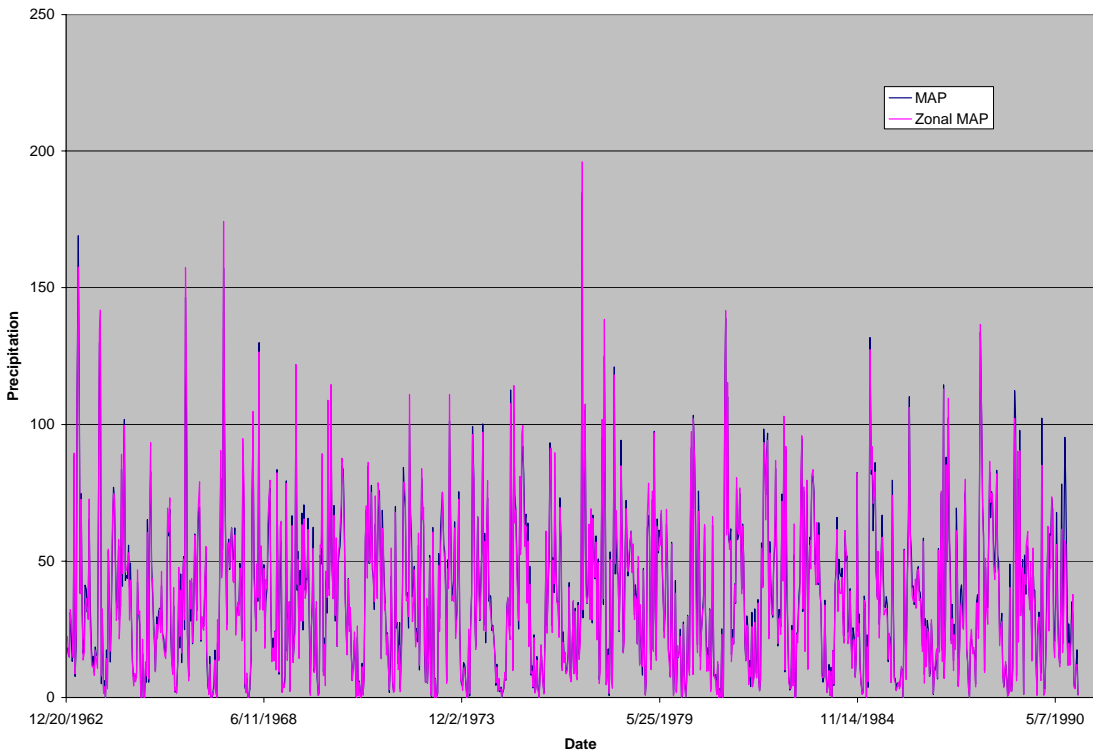


Figure 4.9: 10-day MAP and Zonal MAP for Nzoia

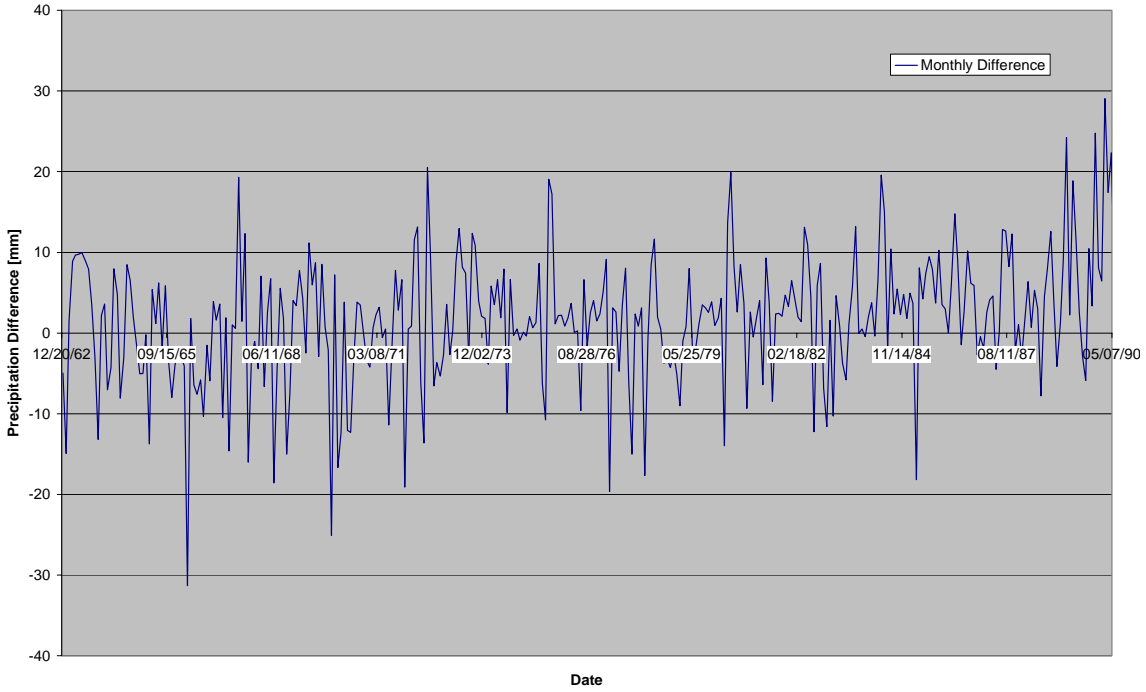


Figure 4.10: Monthly Difference in MAP and Zonal MAP for Nzoia

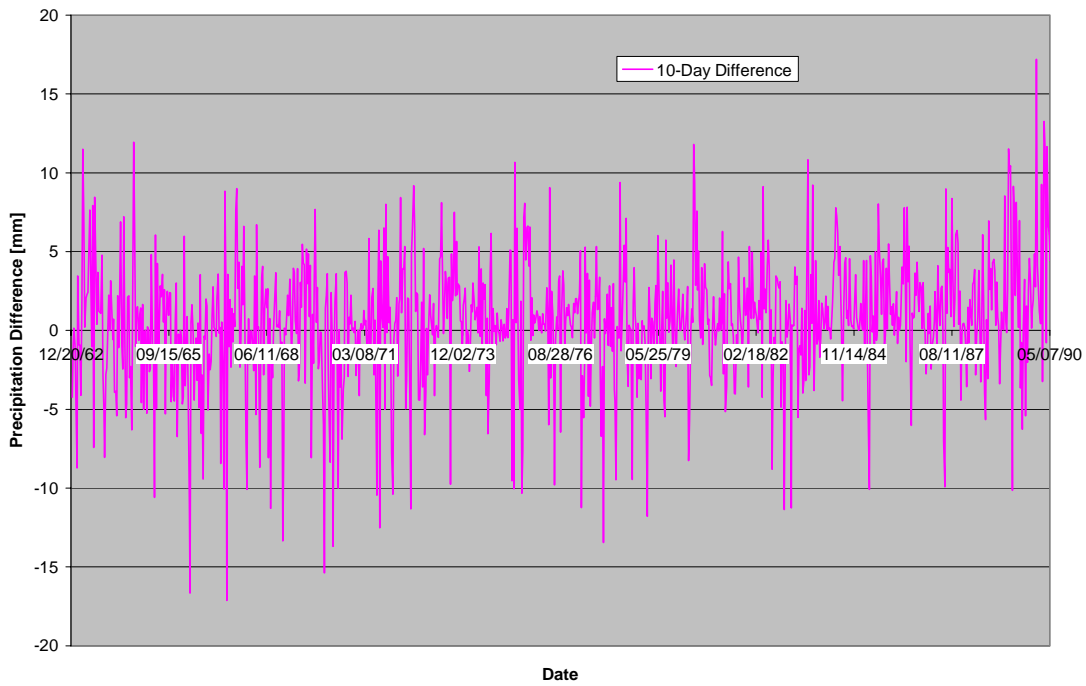


Figure 4.11: 10-day Difference in MAP and Zonal MAP for Nzoia

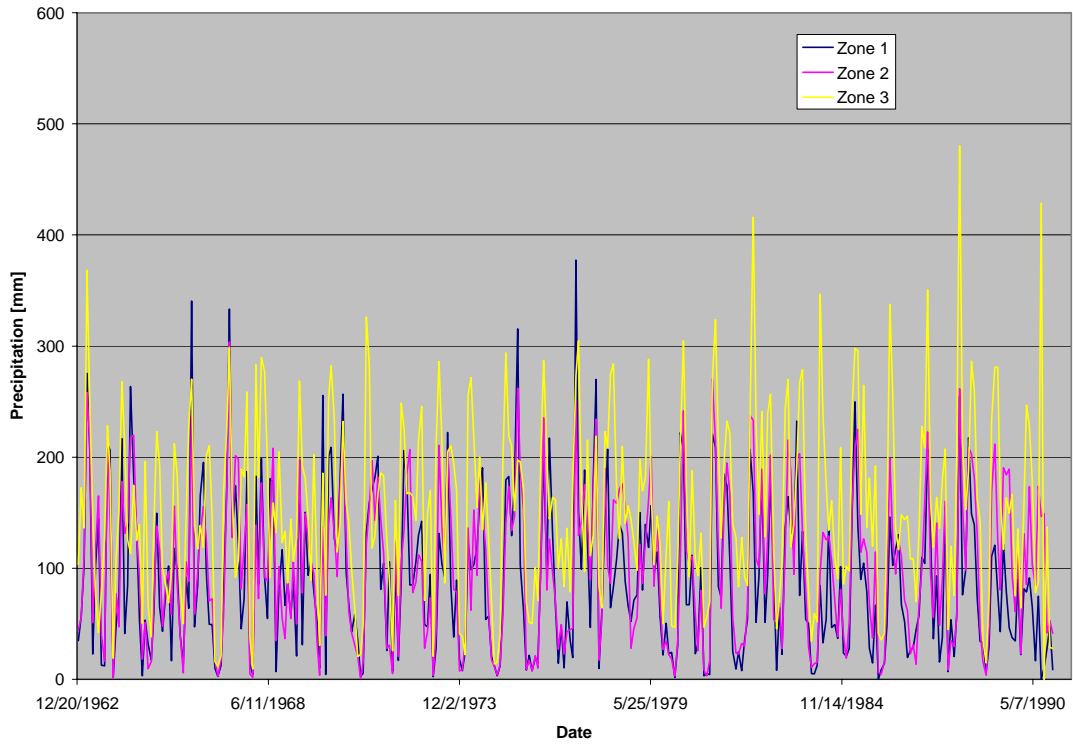


Figure 4.12: Monthly Precipitation per Zone for Nzoia

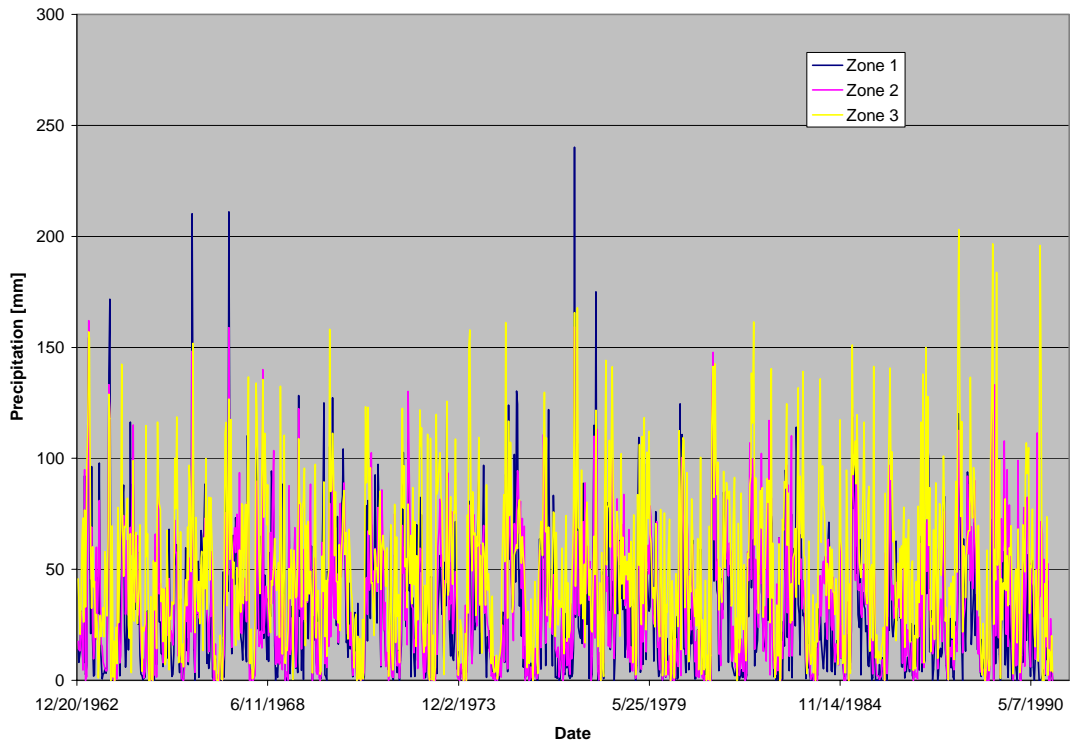


Figure 4.13: 10-Day Precipitation per Zone for Nzoia

4.2 Temperature Estimation

Temperature is the second meteorological parameter used in hydrologic modeling as part of the watershed evapotranspiration estimation. The process of developing the model temperature input is described next.

The Nile DST data base module was used to identify all ground monitoring stations within the Nzoia basin. The station records were plotted to identify those stations containing both maximum and minimum temperature data. In the case of Nzoia, there were five stations within the basin with periods of relatively complete records. While the five stations were collectively representative of all major climate zones, there were many periods for which only one or two stations contained data. This creates a problem in that the average basin temperature calculated based on a time-varying subset of stations is not indicative of the true basin average. As a result, directly applying an inverse distance weighting scheme to develop an average basin temperature series yields temperature variations much greater than the variability observed in any individual station series. Further, the available records did not span a long enough period for calibration of the hydrologic model.

In view of the above, typical climatologies, on monthly as well as 10-day bases, were developed for each of three climate zones from all years of available data within each zone. The zonal climatologies were weighted, with respect to their relative areas, and averaged to produce a typical climatology for the entire basin. The monthly and 10-day climatologies were then used to generate a mean basin temperature series for the desired time period. While such an approach does not exhibit inter-annual variability, it does represent a typical annual series developed from all years of available data. Furthermore, the zonal approach prevents the overestimation of temperature variability seen in the direct application of inverse distance weighting.

The model uses temperature data in the form of potential evapotranspiration (PET). Because the PET function is a non-linear function of temperature, the minimum and maximum temperature series were not directly averaged. Instead, PET was calculated for the maximum and minimum temperature series separately, and the average PET was then calculated.

4.3 Streamflow Record

The last piece of information necessary to develop the hydrologic model pertains to watershed streamflow. The process to develop a reliable streamflow record is described next.

Station Selection

In the case of Nzoia, gage height data was the limiting factor for selecting a time and length for model calibration. Data from several gaging stations were analyzed in an effort to select the station with the longest continuous data set. Three stations were selected for extensive analysis using the Nile DST data visualization tool on the basis of location, length of record, and

completeness of record. While gage height data is provided in the database on a sub-daily basis in most cases, actual flow values are limited, with as few as five to seven values per year.

The three stations selected for further consideration were:

Station 60079, nearest the basin outlet;

Station 60077, upstream of station 60079 along the main branch;

And station 60082, also upstream of station 60079, lying on a tributary just above the confluence with the main branch and very near station 60077.

Station 60079 was considered first due to its position in the basin. While this station has a relatively long record, 1974 to 1999, the position of significant gaps in the data effectively reduce the continuous and usable portion of the series to less than 11 years. However, this station has potential for extending and filling gaps in data by utilizing the two major upstream stations, 60077 and 60082, that provide the majority of flow to this station. Station 60082 was considered only for the purpose of extending station 60079's series, as the flow along this tributary is minor, less than 10%, compared to flow in the main branch.

Station 60077 was considered for two purposes. First, this station was used in conjunction with station 60082 in an attempt to relate upstream flow to the downstream flow at station 60079. In addition, because this station is along the main branch and still quite low in the basin, it was considered as a primary station for model calibration. The available record for this station is approximately 1963 to 1999. However, as with station 60079, significant gaps in the record greatly reduce the useful portion of the series.

Rating Curves

The forms of the rating curves for each of the three stations were not available. Therefore, it was necessary to develop rating curves from the limited number of provided flow values before proceeding into the potential calibration scenarios mentioned above.

Flow values were plotted against corresponding gage heights. In addition, flow values were plotted on a secondary axis along with a plot of gage height versus time. Clearly erroneous points, as seen from both plots, were removed from further analysis. Because single flows were provided for given dates, while gage height was provided twice daily, flows could be plotted against the highest, lowest or average daily gage height. All three scenarios were explored. Further, hydrograph analysis was used to determine which flows were representative of a rising or falling hydrograph limb.

Station 60079

Forty-two flow values were available for development of the station 60079 rating curve. Four erroneous values were removed, leaving 38 for the final curve regression. Analysis of rising and falling limb flows did not reveal any clear rating curve hysteresis; therefore, a single curve was fit to the data, as shown in the figure below.

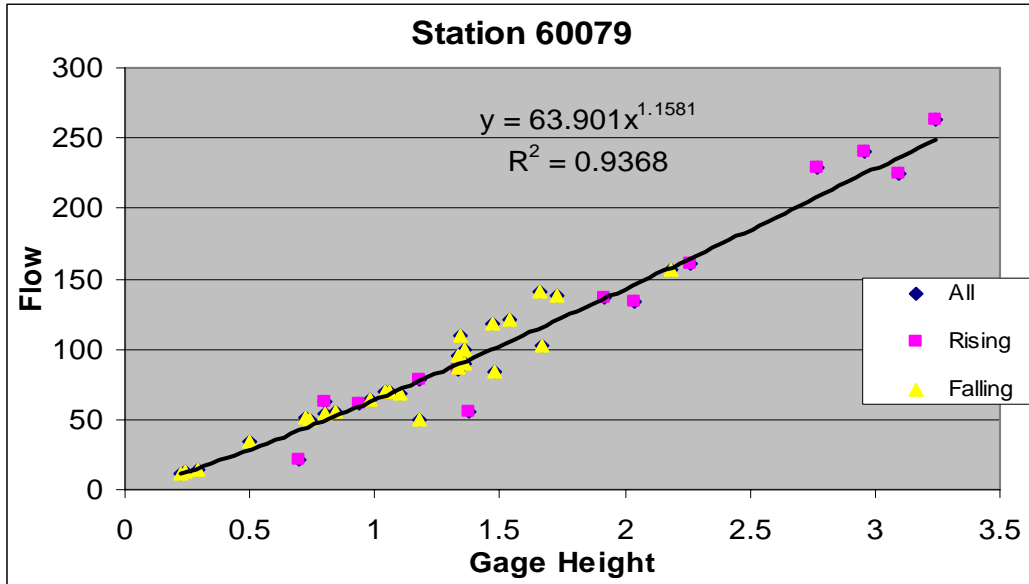


Figure 4.14: Rating Curve for Station 60079

Station 60082

Seventy flow values were available for station 60082 rating curve development. Two points were rejected based on graphical analysis. Again, no clear hysteresis was found, and a single-limbed rating curve was developed.

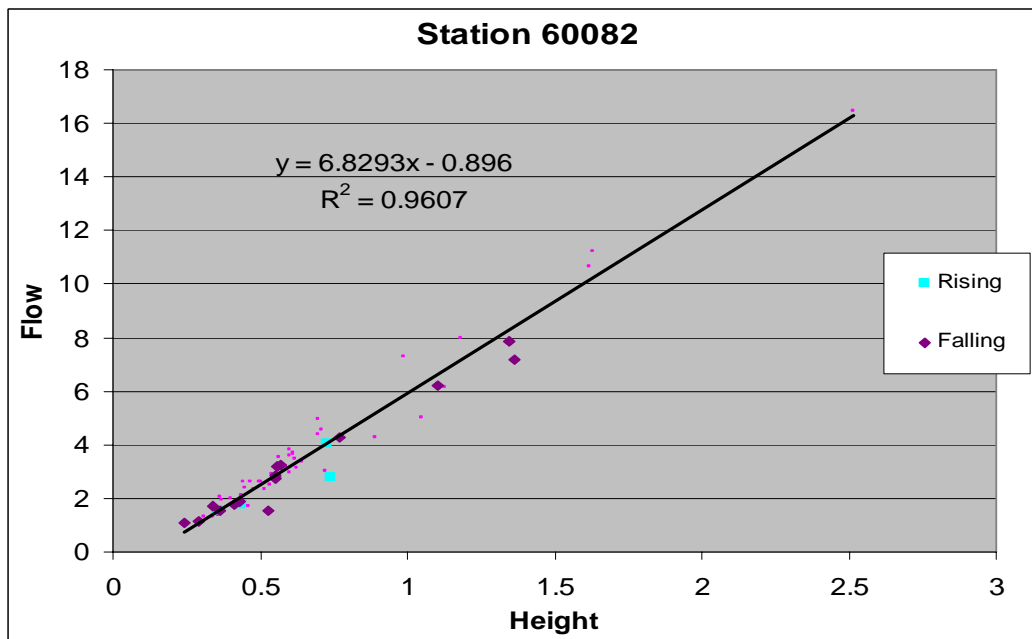


Figure 4.15: Rating Curve for Station 60082

Station 60077

Two different data sets are available for station 60077. Sixty five flow values, from 1964 to early 1988, were provided as part of the Nile DST database. A much more extensive set of flow values, approximately 2900 values spanning 1970 to 1978, are available from the Lake Victoria Decision Support System (LVDSS) database. Both sets were plotted and analyzed in an effort to fit the best possible rating curve to the data. However, the majority of LVDSS flows appear to have been generated from existing, yet undocumented, rating curves. As can be seen in the figure below, two distinct trends appear in this data set. Most flow values from 1970 fall along one curve, while data from 1971 through 1978 appear to fall along another, completely different, curve. The existence of so many points that have clearly been generated based on existing rating curves will overwhelm any attempt to develop another curve from all of the data. Further, it is unknown whether direct use of these two trends is appropriate, as there remains scatter in the data set. There is no way to know whether the scatter results from plotting daily flows against lowest, highest or average daily gage heights; a combination of generated and actually observed flow data; and/or erroneous points in the database.

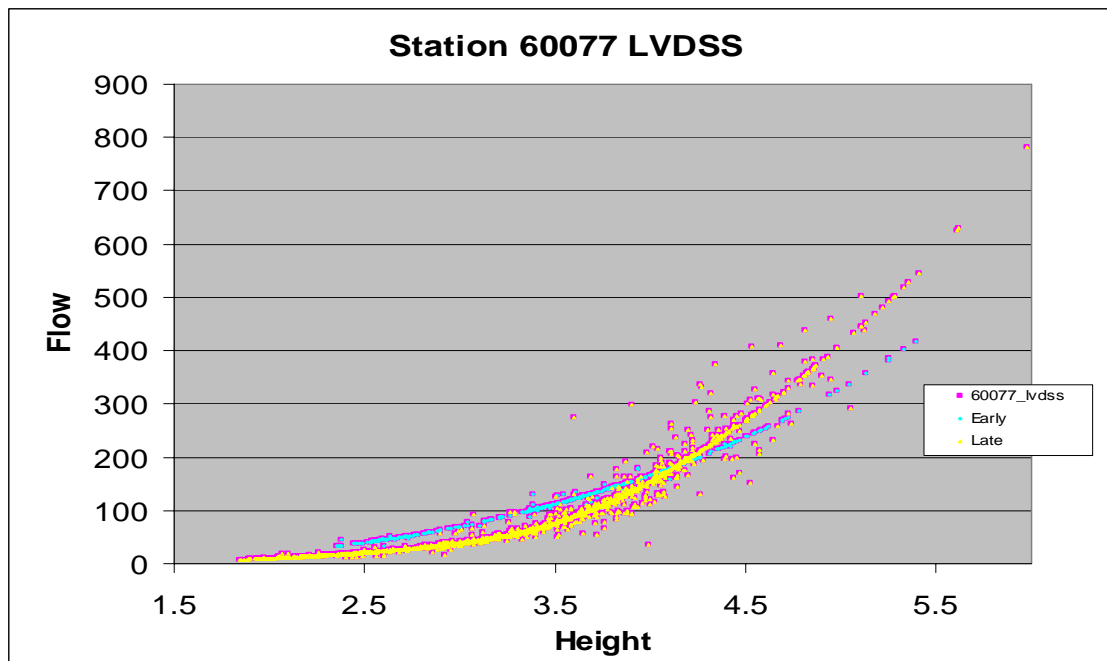


Figure 4.16: Flow-Height Data Relationship; Station 60077; Hydromet database

While there are fewer flow values provided through the Nile DST database, they do not exhibit the same artificial behavior seen in the LVDSS flows. Rising and falling limb flows within this data set were analyzed for possible differences in rising and falling hydrograph rating curves. Again, there was no discernable hysteresis in the data.

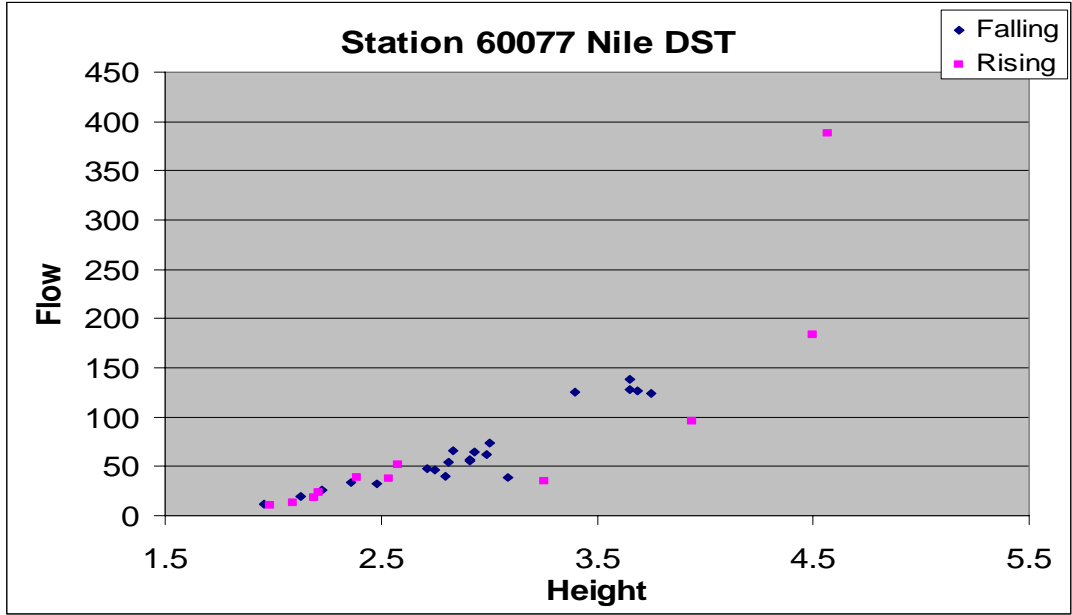


Figure 4.17: Hydrograph Analysis for Station 60077

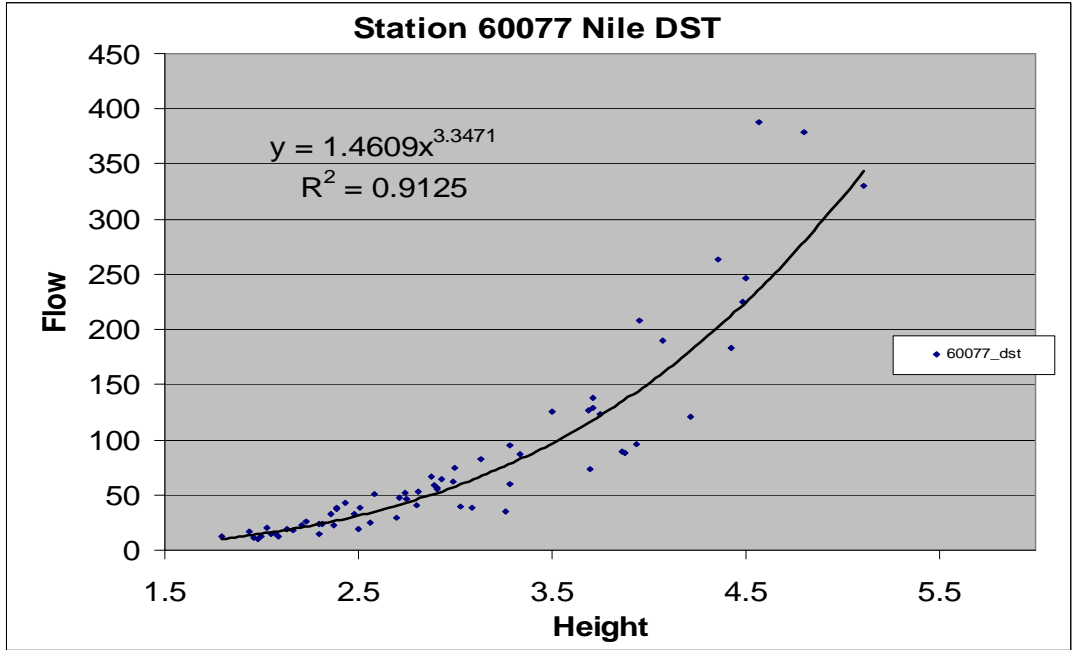


Figure 4.18: Rating Curve for Station 60077; Nile DST database less Hydromet data

It was not immediately clear which set of data to rely on for developing a rating curve. Two scenarios were constructed and evaluated to aid in selecting a rating curve. In the first scenario, the station 60082 rating curve was applied to gage height data to generate a corresponding flow series. Then the two time-period specific rating curves based on the LVDSS data for station

60077 were applied to gage height data to generate the flow series for that station. The two series were summed and plotted along with the series from station 60079 downstream.

In the second scenario, flows at stations 60082 and 60079 were generated identically. However, station 60077 flows were generated using the rating curve derived from the Nile DST flow data. Again, the summed flows from stations 60082 and 60077 were compared with the downstream flows at station 60079.

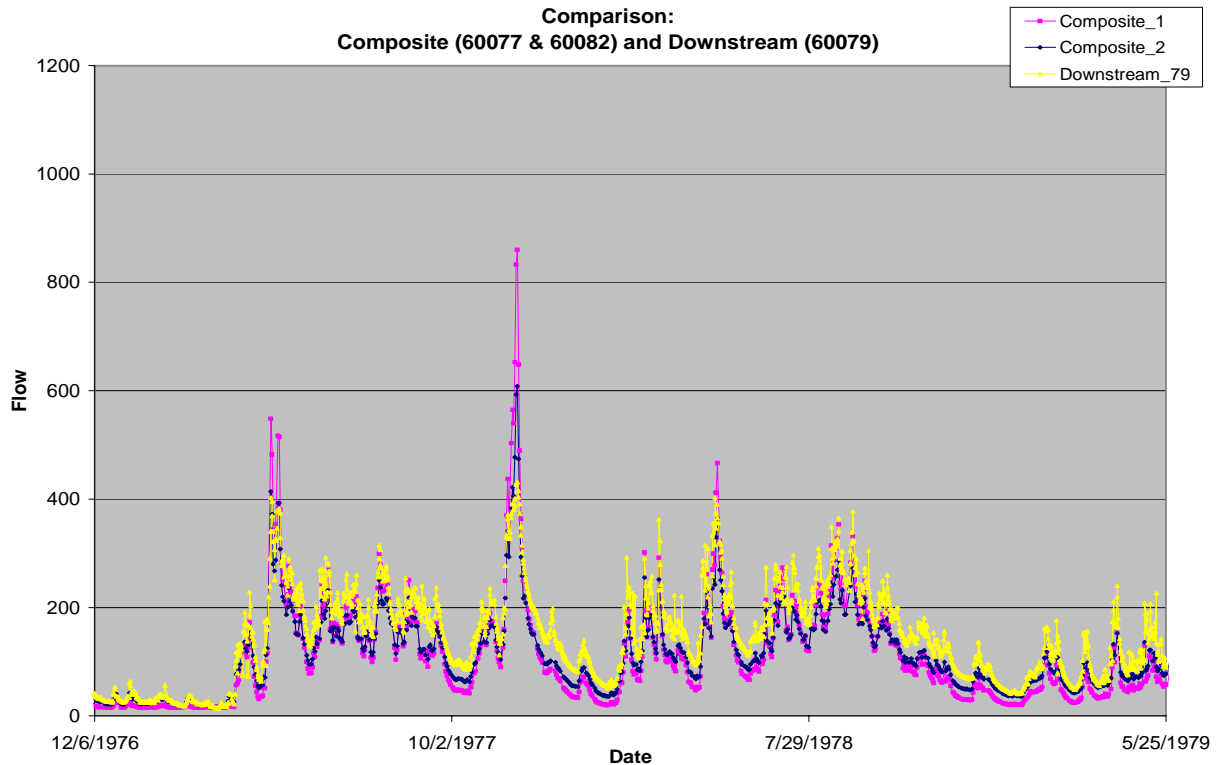


Figure 4.19: Comparison of Station 60079 with Combined Flows from Stations 60077 and 60082

As the figure above illustrates, scenario two (blue series), which utilizes the Nile DST-derived rating curve, appears to yield a better fit with the downstream flow series. This additional information led to a decision to use the Nile DST-derived rating curve at station 60077.

Streamflow Data Analysis and Extension

Station 60077 was selected for model calibration based on the duration and completeness of the data. After some investigation, the longest continuous period available is between 1974 and 1985. However, within this period there are data gaps of two to 10 days. Because the model is run on a 10-day time step and flow is a cumulative value rather than an average over the period, the missing data could not be ignored. Several options were considered for addressing this problem.

First, the flow data was plotted and analyzed local to all areas of missing data. If the missing values were single days within a clear hydrograph trend it may be possible to evaluate the approximate magnitude of the missing value. In reality, several areas contained multiple consecutive missing values, and these values could not easily be estimated based upon the local hydrograph trends.

It was known from previous analysis that flows at station 60077 are closely related to flows at the downstream station 60079. Regression analysis on both daily and 10-day flows were conducted to assess the correlation between the two stations. In the case of daily flow both single and multiple linear regression yielded correlation coefficients in the mid 90% range. While this is good correlation, use of the regression model for filling in missing data was not appropriate as model errors masked the finer scale hydrograph features and thus produced inconsistent values.

Both linear and non-linear regression between the stations was carried out on 10-day flows. The linear regression had a 0.95 correlation coefficient, while the non-linear regression yielded a correlation coefficient of 0.98. Because the two series have a very high linear correlation, it was possible to relate an incomplete 10-day interval at station 60077 to the ratio of the corresponding complete 10-day interval and an artificially incomplete interval of station 60079. More explicitly, the complete 10-day intervals at the downstream station were altered to remove the same daily values as were missing at station 60077. The ratio of the complete 10-day average and the incomplete 10-day average for station 60079 was then applied to the incomplete average at station 60077 to yield its estimated complete 10-day average. This technique was applied throughout the series and plotted and checked for consistency. The results fit very well within the local hydrograph behavior.

The non-linear regression, which yielded an even higher 0.98 correlation coefficient, was used to extend the period of record at station 60077 based on a complete flow record at station 60079. Application of this regression model was checked using periods where flows exist at both stations. The error in estimated station 60077 flows was generally found to be well below 10%. Use of this relationship effectively extended the available calibration data to the end of 1987.

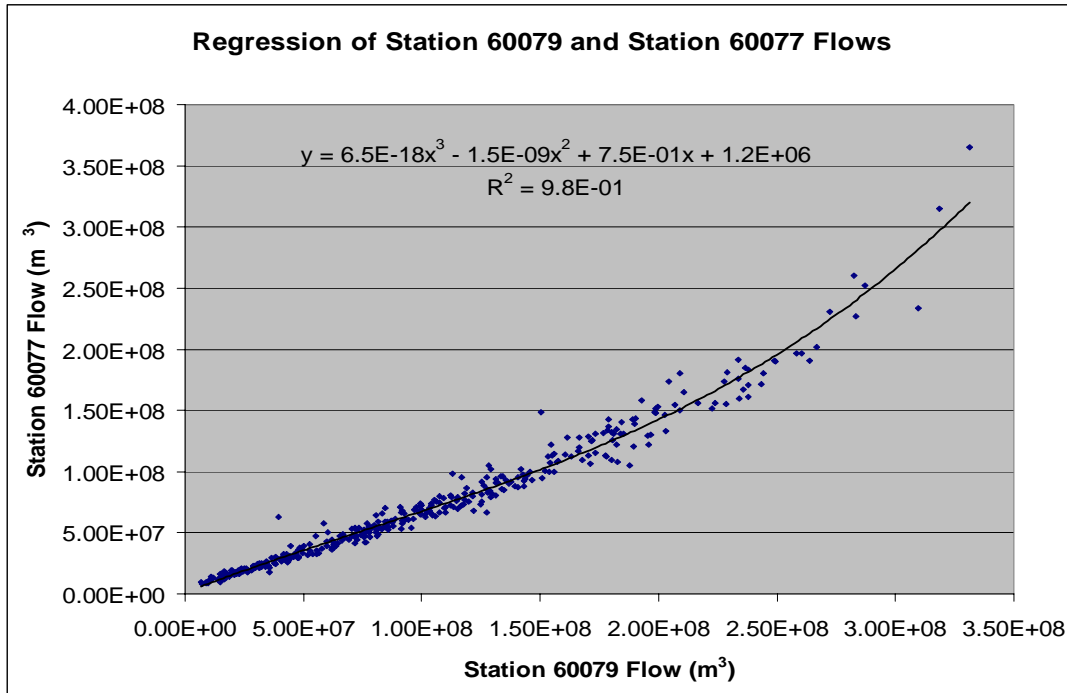


Figure 4.20: Regression Analysis for Stations 60079 and 60077

4.4 Model Calibration and Results

The total continuous data series for precipitation, temperature-based evapotranspiration and discharge available for model calibration extends from the beginning of 1974 to the end of 1987. Of the 14 years, 10 were used directly in calibration, with the remaining four used to verify model results. The model-generated flows for the verification period, 1984 to 1987, are shown in the figure below. (Simulated in red, observed in blue)



Figure 4.21: Model Verification (1984 - 1987)

After verification, the calibrated model was used to simulate the entire 14 years of streamflow and compared with observed flows. Simulated and observed flows are shown and compared in the figure below.

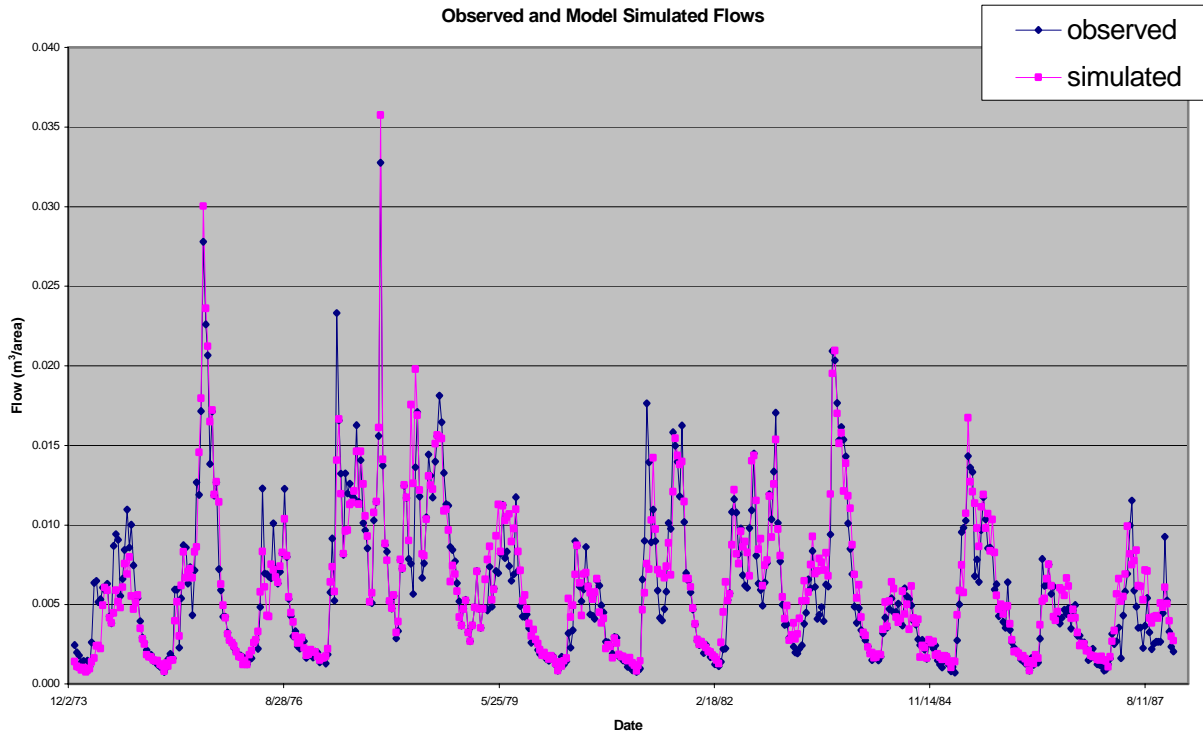


Figure 4.22: Simulated versus Observed Flows (1974 – 1987)

As can be seen, the model performed quite well, especially given only 10 years for calibration. While the magnitude of a few peaks is off slightly, the overall hydrograph trends are captured very well, especially in the recession limbs and mid-range peaks. As more data becomes available or is developed to become appropriate for calibration, model results will continue to improve. Even with the present available sequence, it may be possible to refine these results somewhat by developing the potential evapotranspiration sequence first at the climate zone levels and then averaging over the basin. This could have minor effects on the timing and magnitude of streamflow response depending on conditions in both the mountainous and lowland regions.

The data analysis and model calibration process just outlined exemplifies the critical role of country hydrologists in developing the hydrologic model. Close familiarity with the data will undoubtedly lead to better model calibration. It should also be noted that compared with other basins, Nzoia data records, despite their gaps, are relatively long. Efforts should be made to develop such records for other Nile Basin watersheds.

5. Kagera Watershed

The Kagera Watershed covers portions of four countries, Burundi, Rwanda, Tanzania and Uganda. The river flow into Lake Victoria and covers approximately 58,000 km² (Figure 5.1).

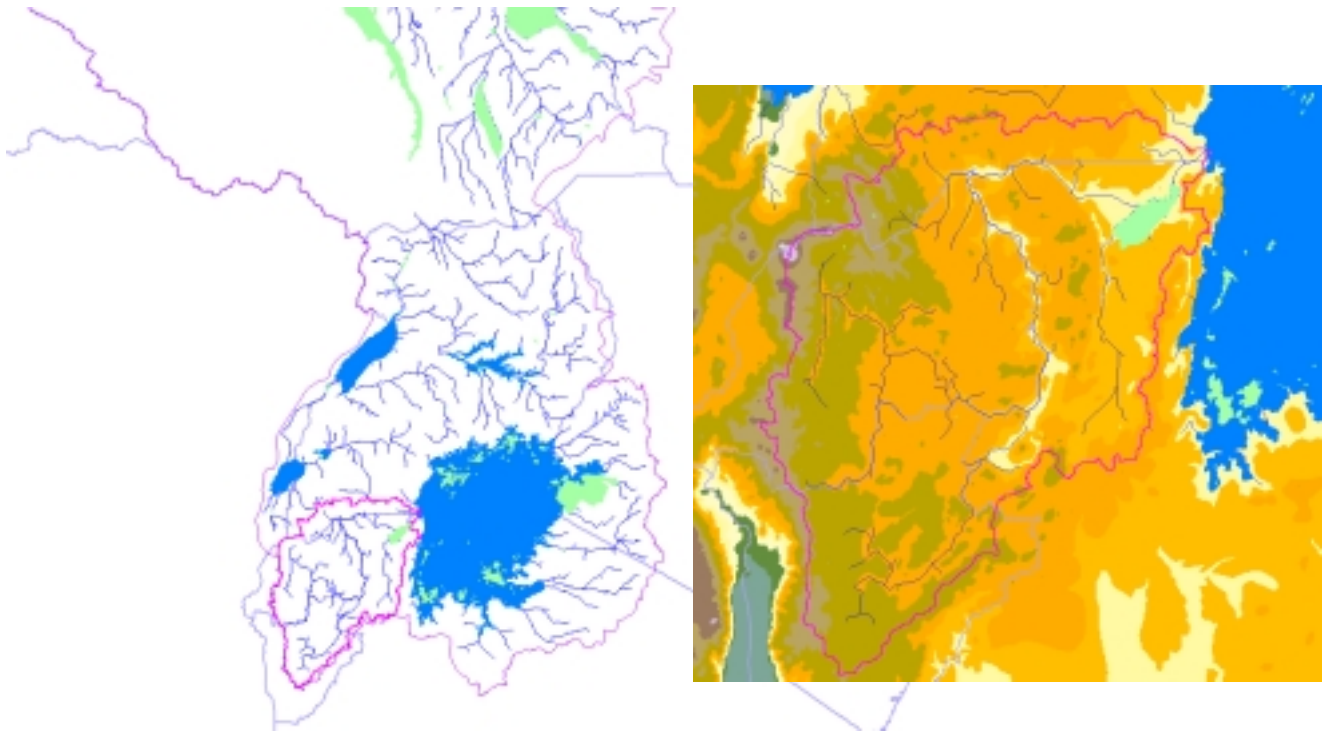


Figure 5.1: Kagera Watershed

5.1 Precipitation Estimation

The precipitation input file for the Kagera basin was developed using two methods. The first method is a mean areal precipitation (MAP) time series developed for the whole basin using a gridded inverse distance weighting (IDW) method. The second method builds upon the first by adding a normal weighted component within the IDW method. The normal weight component adds a grid elevation consideration into the calculation.

Station Selection

The stations, to be used in both methods, were selected by using the features available in the database, map and chart components of the Nile DST. The watershed for the Kagera basin was selected in the map view and the 'selected by theme' tool was used to select all precipitation stations within the watershed. Once selected the stations were charted using a yearly time interval and a statistic of count (Figure 5.2). The stations with a significant number of years of records (generally more than ten) and yearly counts of 90% (approximately 330) were selected.

Unfortunately, a large swath through the middle of the watershed has no stations with significant records (Figure 5.3).

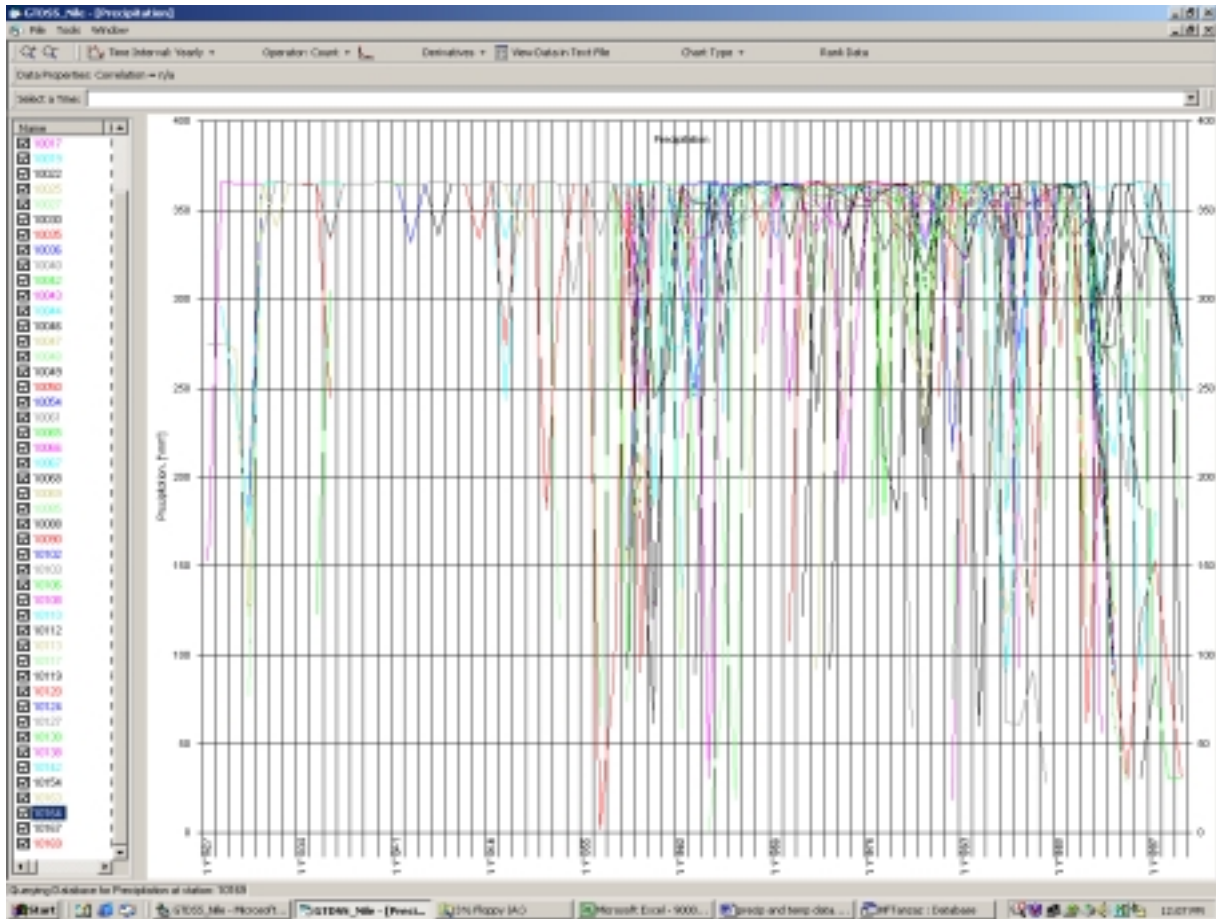


Figure 5.2: Precipitation Data in a Yearly Count Chart

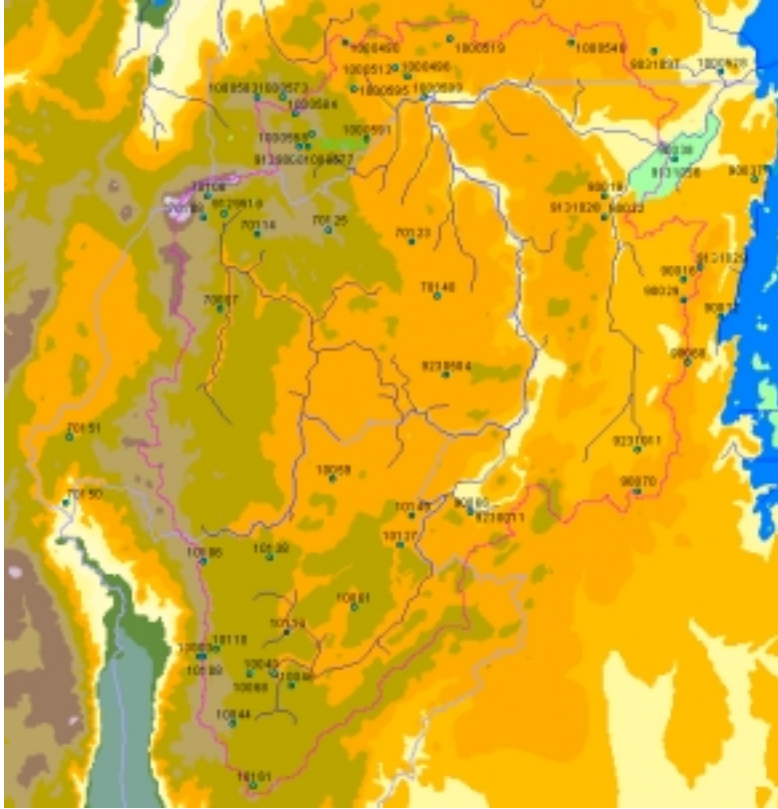


Figure 5.3: Selected Precipitation Stations

MAP Estimation

The mean areal precipitation (MAP) was calculated for the watershed using the same IDW method described previously.

Normal Weight MAP Estimation

The normal weight MAP method uses a normally weighted value of precipitation based on elevation considerations within the gridded IDW method. The precipitation is modified using the following relationship:

$$P_i = \frac{N_a}{N_i} P_a$$

Where,

P_i = The normal weight precipitation value (mm)

P_a = Station 'a' precipitation record (mm)

N_a = Average Annual Precipitation (AAP) for station 'a'

N_i = Calculated AAP based on the grid centroid elevation

In order to calculate the N_i each grid a functional relationship for AAP versus elevation must be developed. For each selected station an AAP for the years of record was calculated and graphed as a function of station elevation (Figure 5.4).

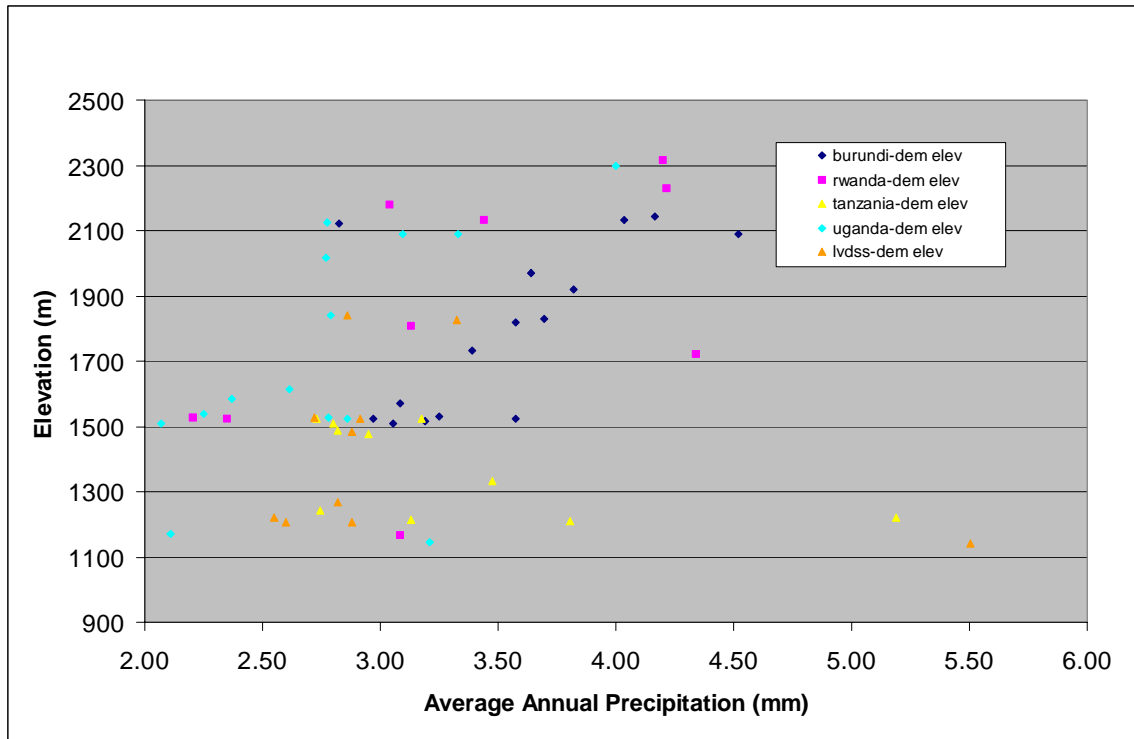


Figure 5.4: Average Annual Precipitation of Selected Stations versus Elevation

In order to analyze the spatial relationships, of the average annual precipitation values were added to the station shape file in ArcView. Using the spatial analysis extension, a 0.1° surface grid was created. The grids were assigned a value using the IDW method with the station AAP values as input (Figure 5.5)

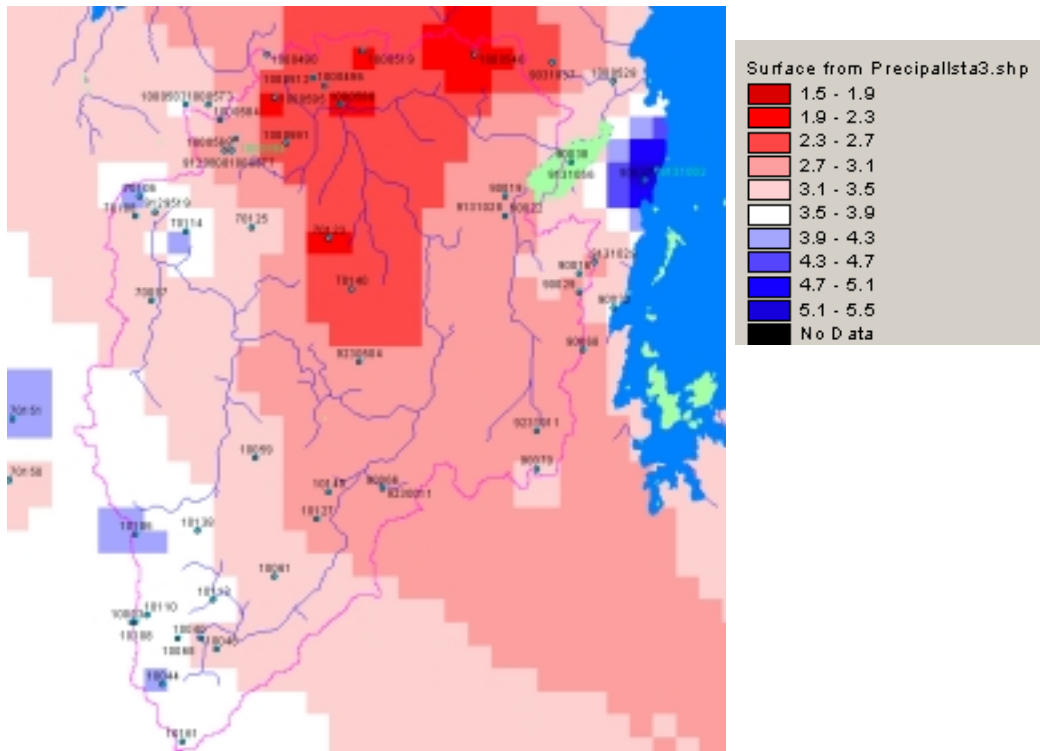


Figure 5.5: Spatial Analysis of Average Annual Precipitation

A few stations demonstrated AAP values significantly different from neighboring stations at similar elevations. These stations were removed from further analysis. The spatial analysis led to the decision to develop an AAP-elevation function for two distinct regions (Figure 5.6).

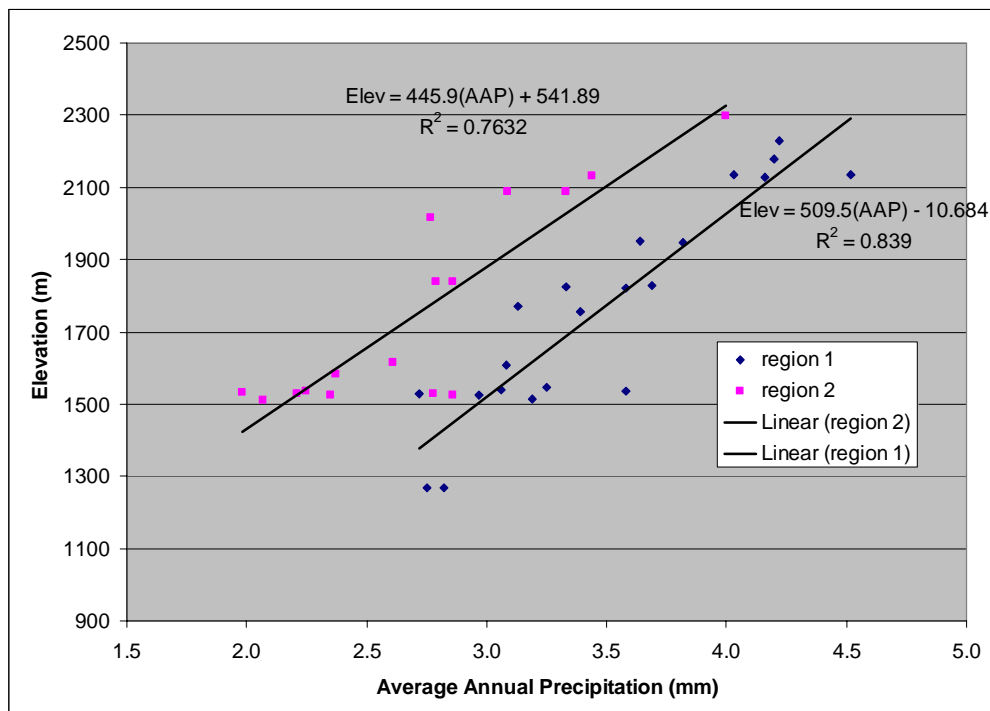


Figure 5.6: Two AAP-elevation functions

Results

The time period for the development of the precipitation input files is based on the stream flow record available for the basin. For the Kagera basin this is from 1960 to 1977. The previously selected stations were used in the calculation of the MAP if the station had 90% of the yearly record for 15 of the 18 years (Figure 5.7).

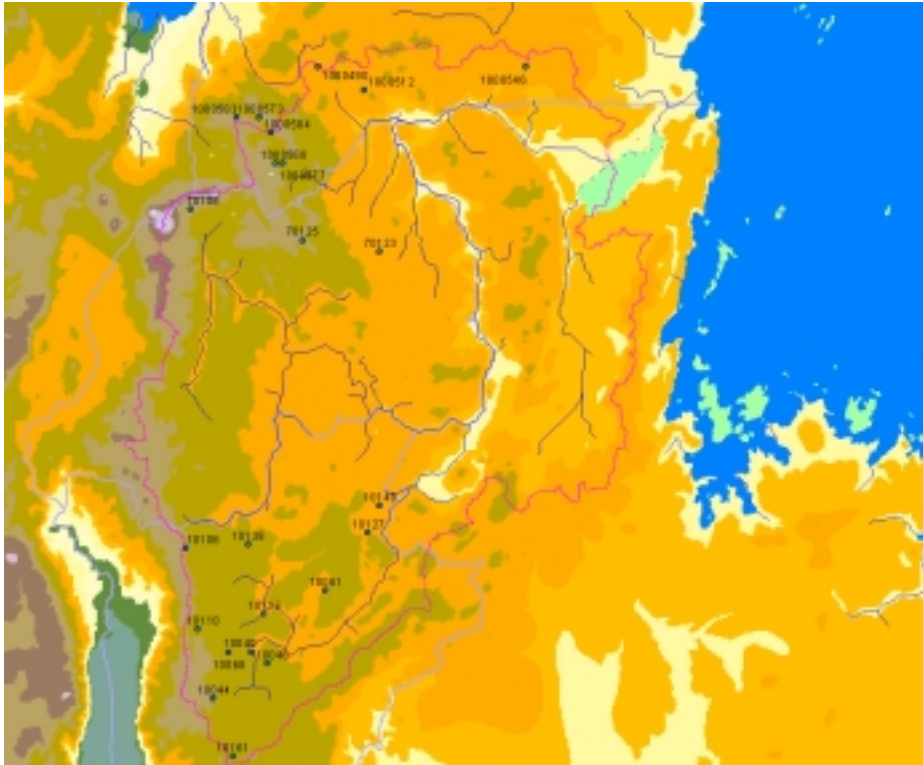


Figure 5.7: Stations used in the calculation of the Mean Areal Precipitation (all three methods)

The MAP, Normal Weight MAP with two AAP-elevation functions and the Normal Weight MAP with one AAP-elevation function were developed for the 1960-1977 time period using the methods previously described. Figure 5.8 and 5.9 are a plot of the monthly and 10-day values respectively for each method. Figure 5.10 and 5.11 chart the difference between the three precipitation series for monthly and 10-day time intervals.

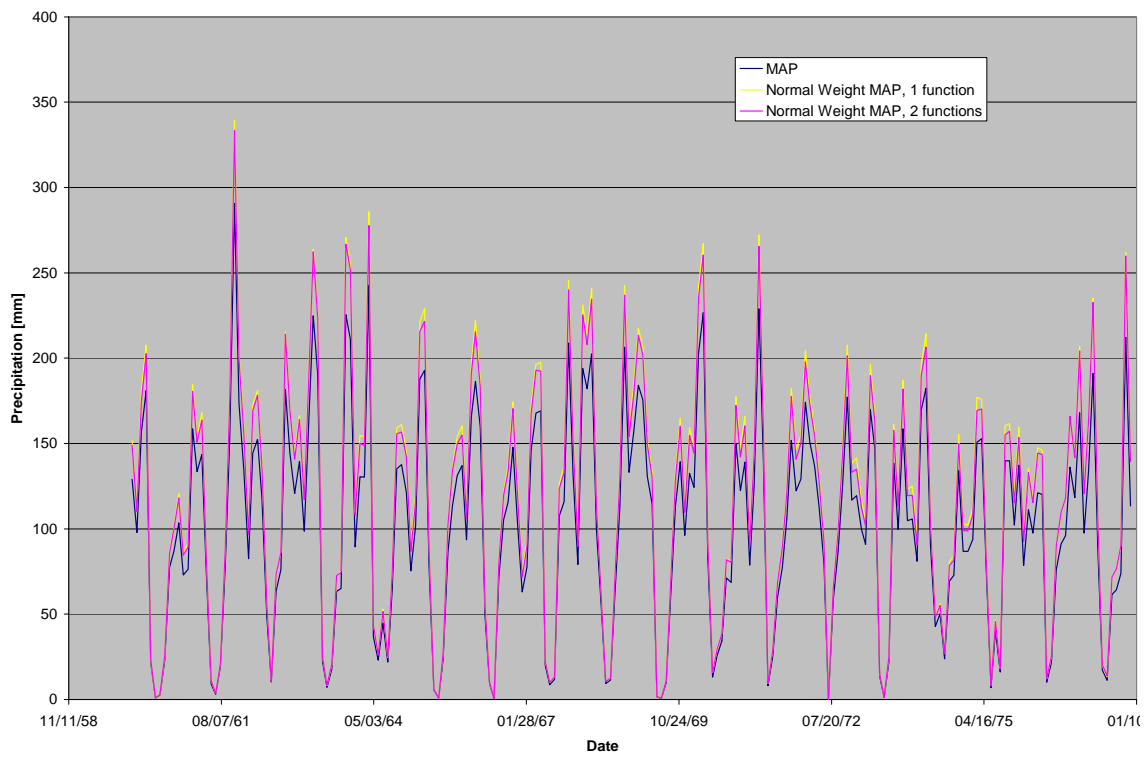


Figure 5.8: Monthly Precipitation for Kagera

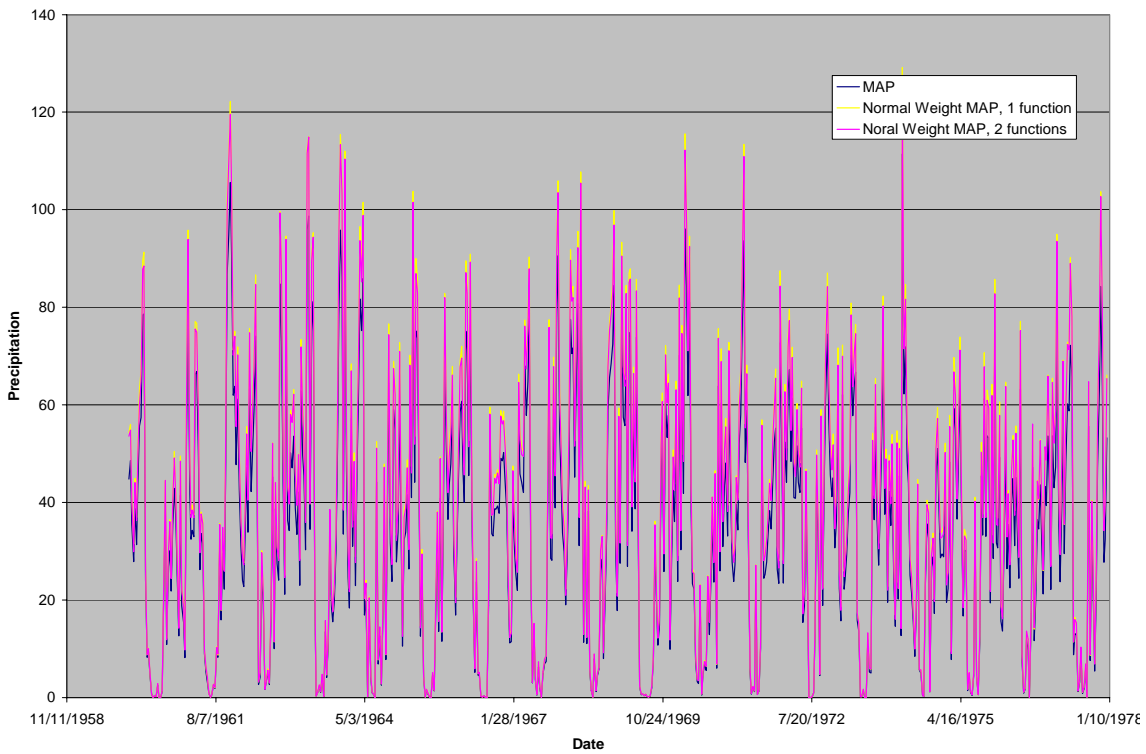


Figure 5.9: 10-day Precipitation for Kagera

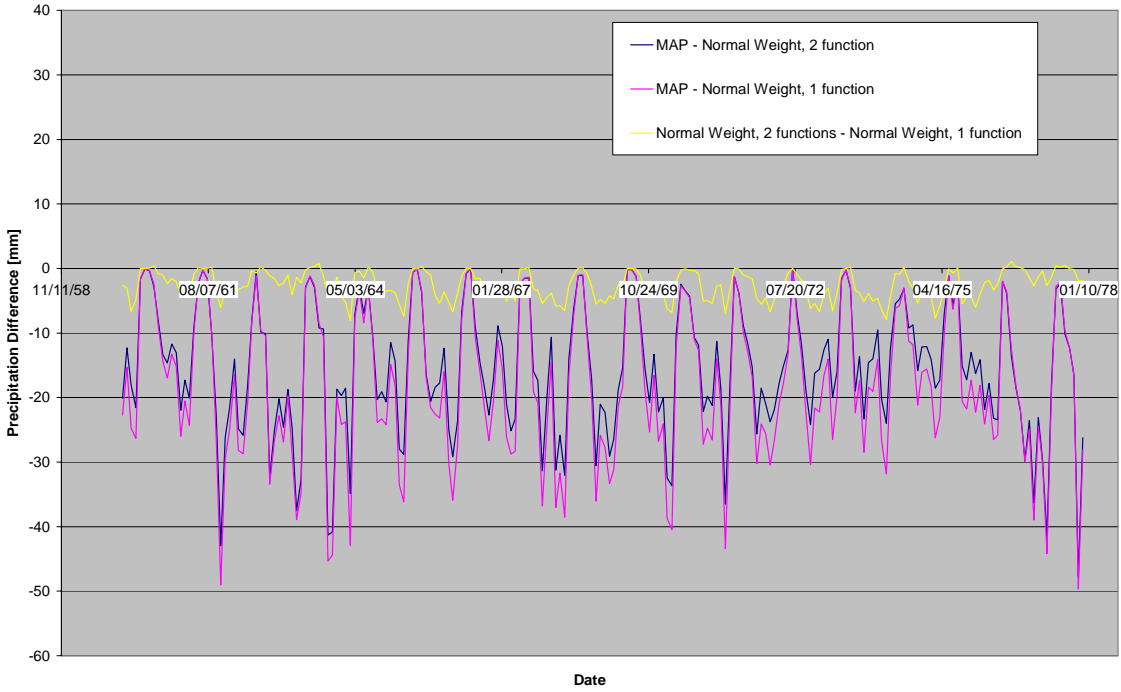


Figure 5.10: Difference in Monthly Precipitation

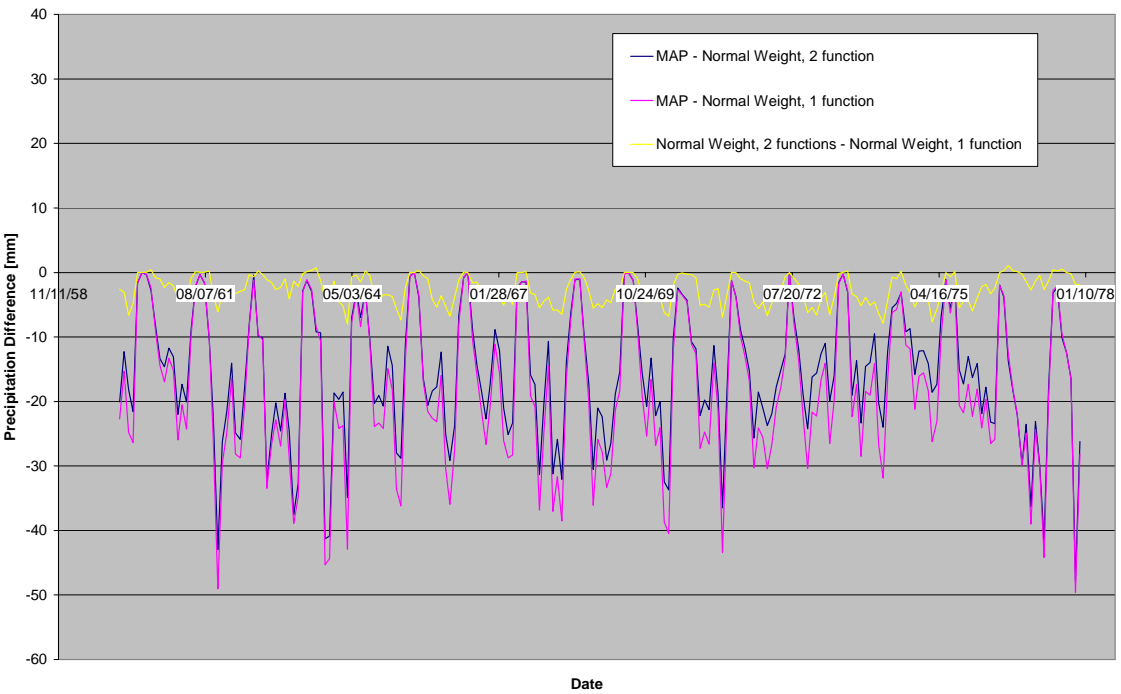


Figure 5.11: Difference in 10-day Precipitation

5.2 Temperature Estimation

The temperature input file for the Kagera Basin was estimated using a method similar to the normal weighted MAP method employed for the calculation of the precipitation input file. Both 10-day and monthly input files were developed. The first step is the selection of stations with adequate temperature records.

Station Selection

Using the same procedure outlined previously for precipitation, the temperature stations were selected (Figure 5.12).

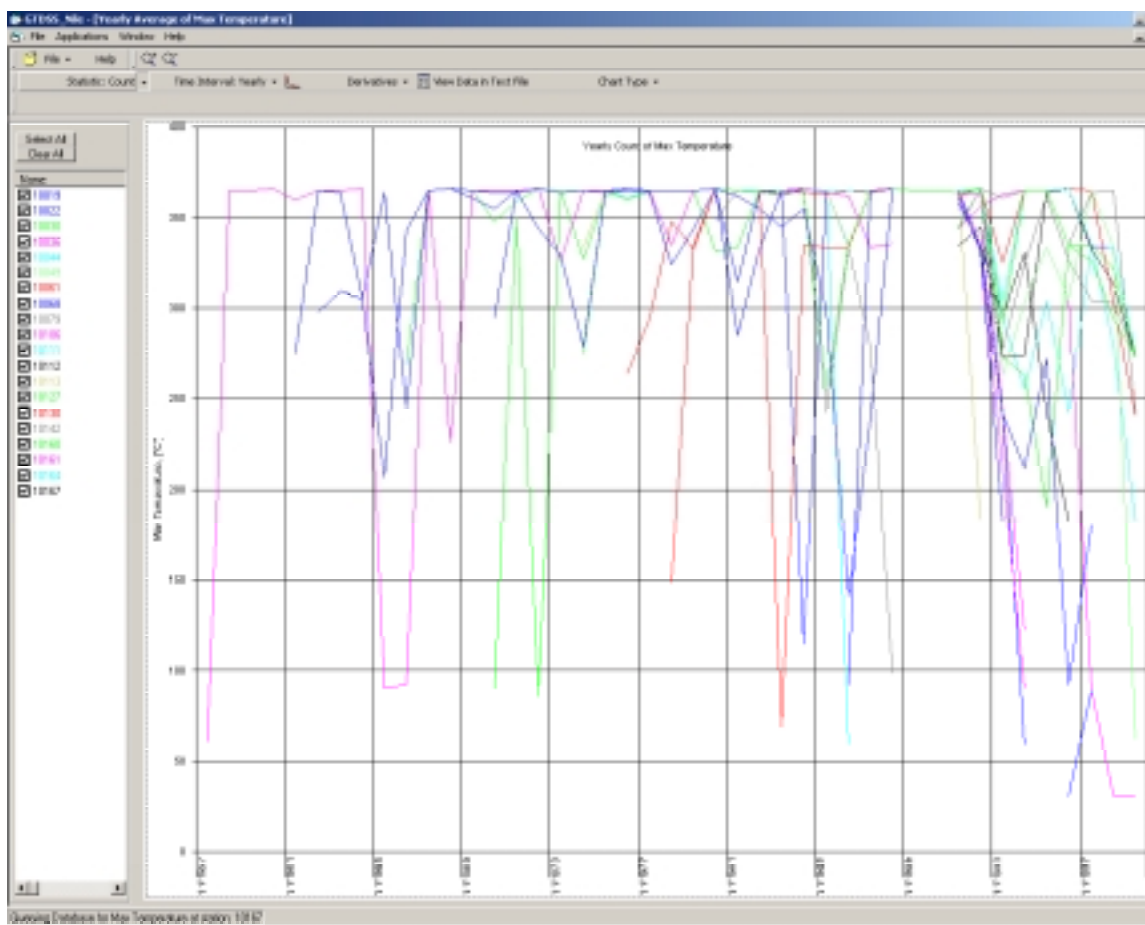


Figure 5.12: Yearly Count Chart of Temperature Data

The majority of the selected stations (Figure 5.13) did not have a temperature records long enough to span the selected model time period (1960-1977). As a result the Nile DST Data analysis component was used to develop a climatology (Figure 5.14), which a monthly average for the record period, for each station. This climatology is used as the station input for the calculation of the temperature input file.

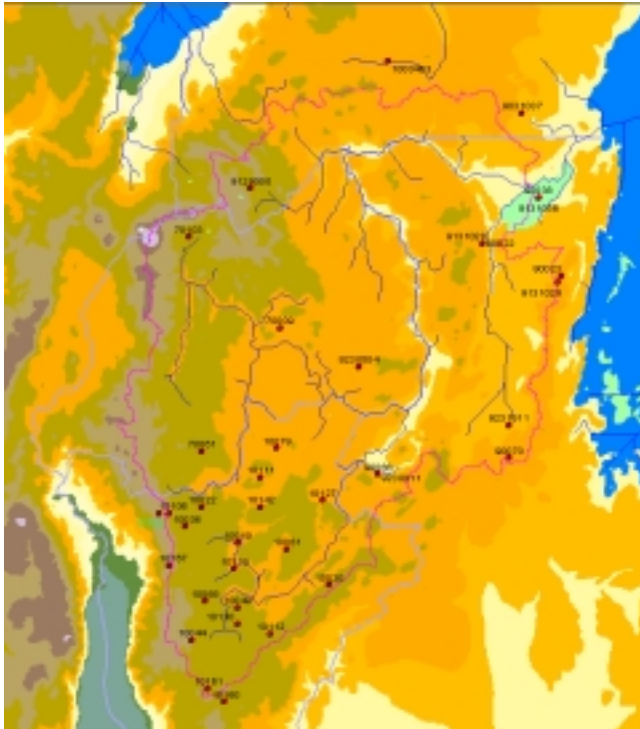


Figure 5.13: Selected Temperature Stations

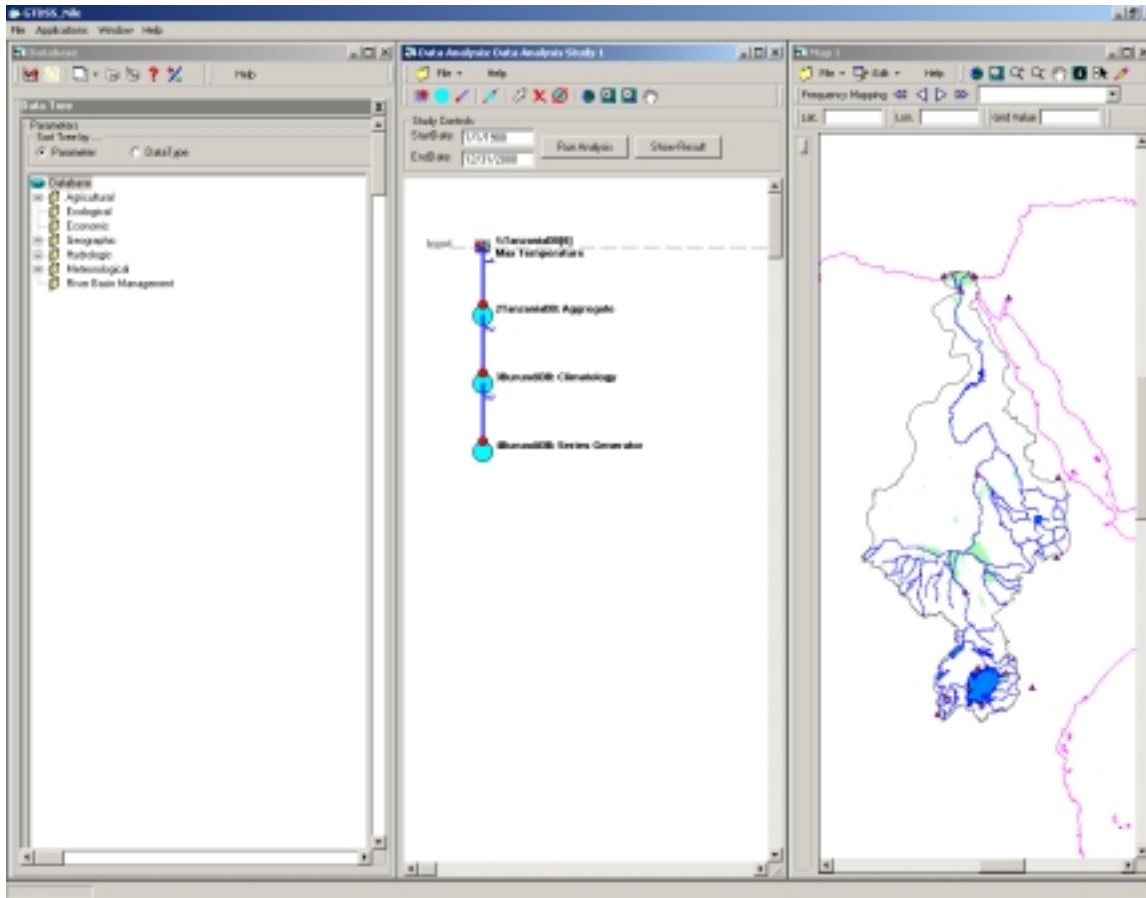


Figure 5.14: Data Analysis Structure used to Generate Climatologies

Both maximum and minimum temperature data were considered during the station selection process. However, several inconsistencies were noted in the minimum temperature files. As a result, the temperature series was developed using the maximum temperature data only.

Temperature Series Development

In order to implement the normal weighted method a temperature-elevation function must be developed. As a result, the annual average temperature (AAT) was plotted against station elevation (Figure 5.15).

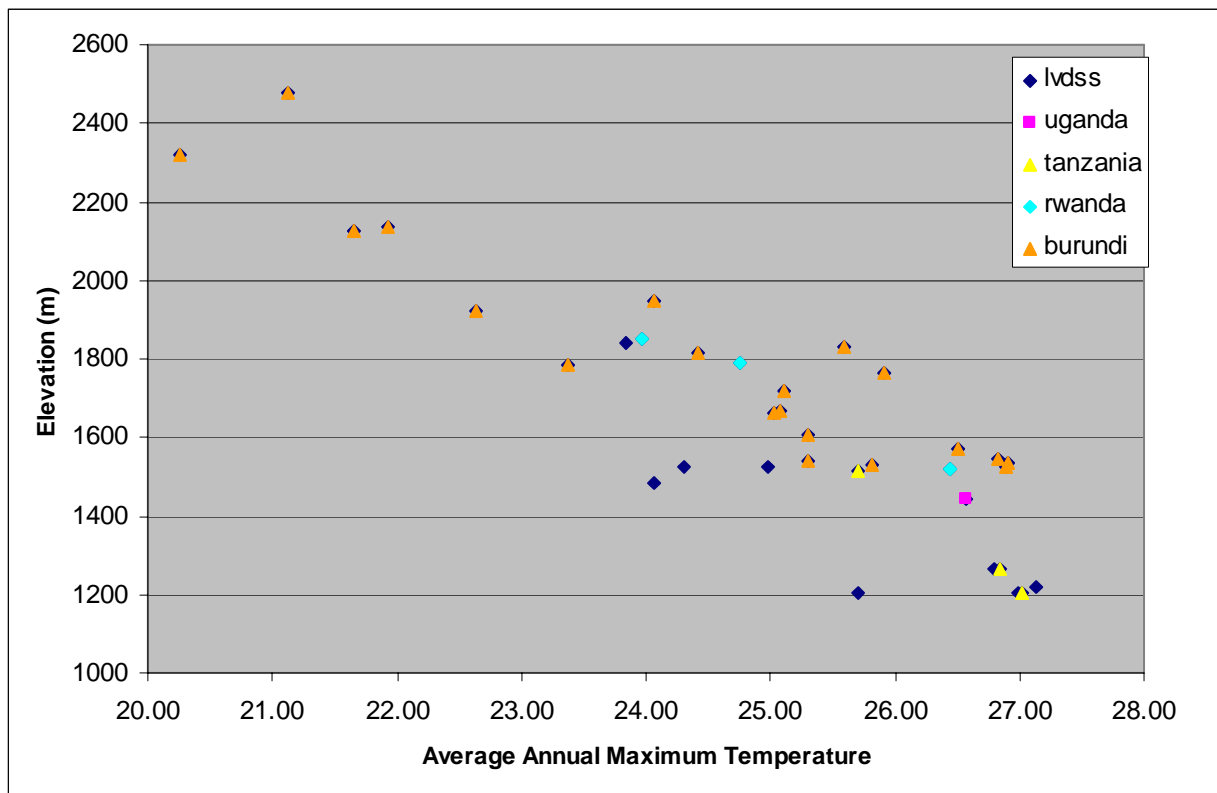


Figure 5.15: Average Annual Temperature versus Elevation

This plot shows a relatively linear decrease in temperature with elevation. The monthly average temperature was also plotted as a function of elevation to determine if the linear relationship was true for monthly data (Figure 5.16). Figure 5.17 shows this plot for only two of the twelve months. The linear pattern remained consistent for the monthly data.

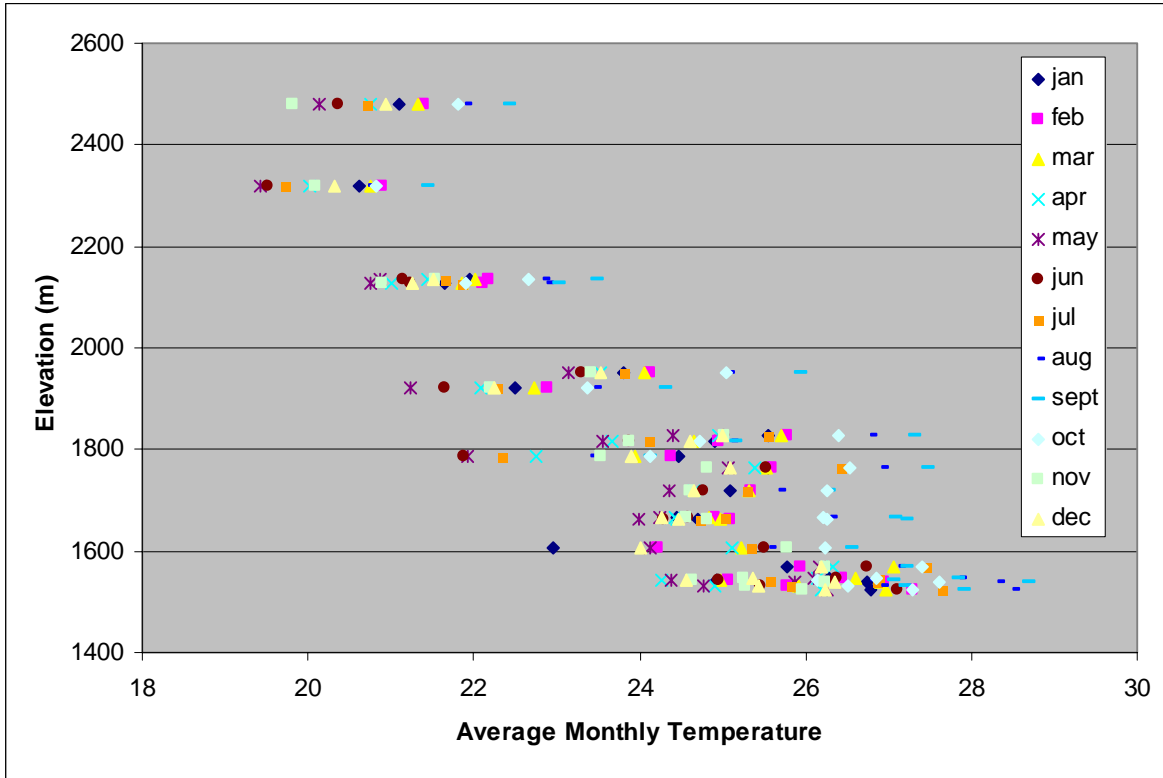


Figure 5.16: Average Monthly Temperature versus Elevation

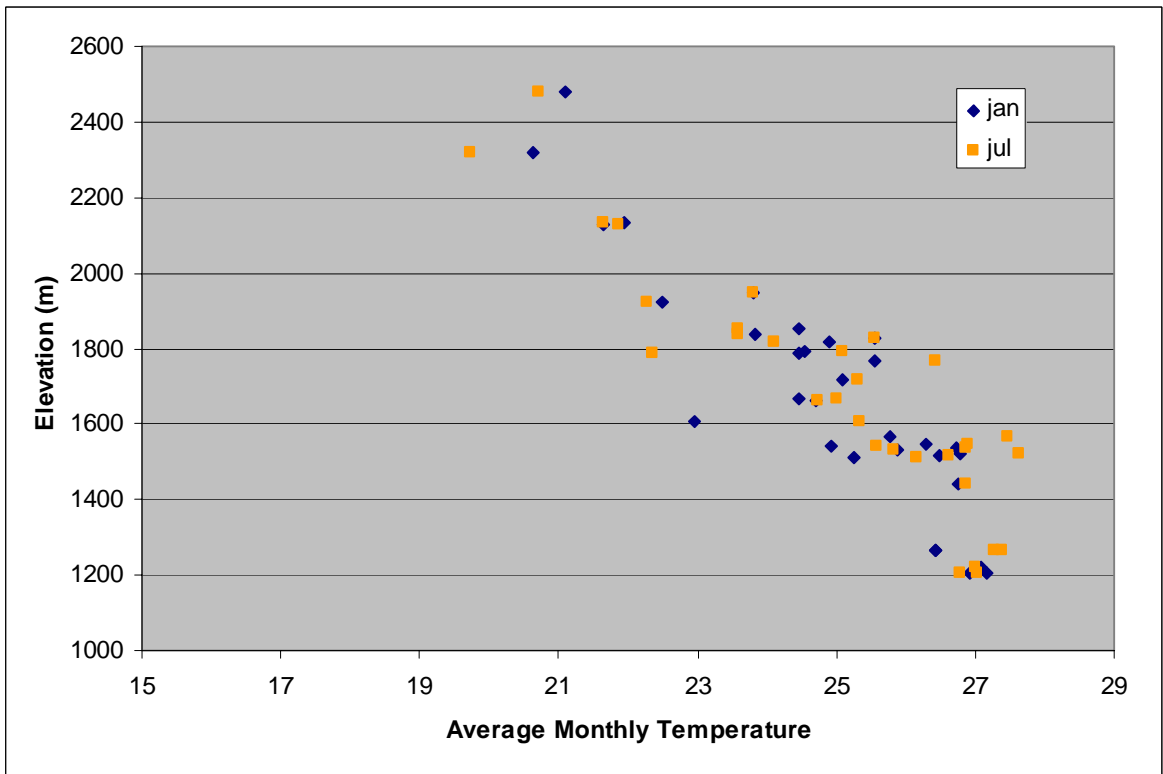


Figure 5.17: Average Monthly Temperature (January and July) versus Elevation

Lastly, the AAT data was analyzed spatially using the IDW method to create a surface from the AAT values for each station (Figure 5.18).

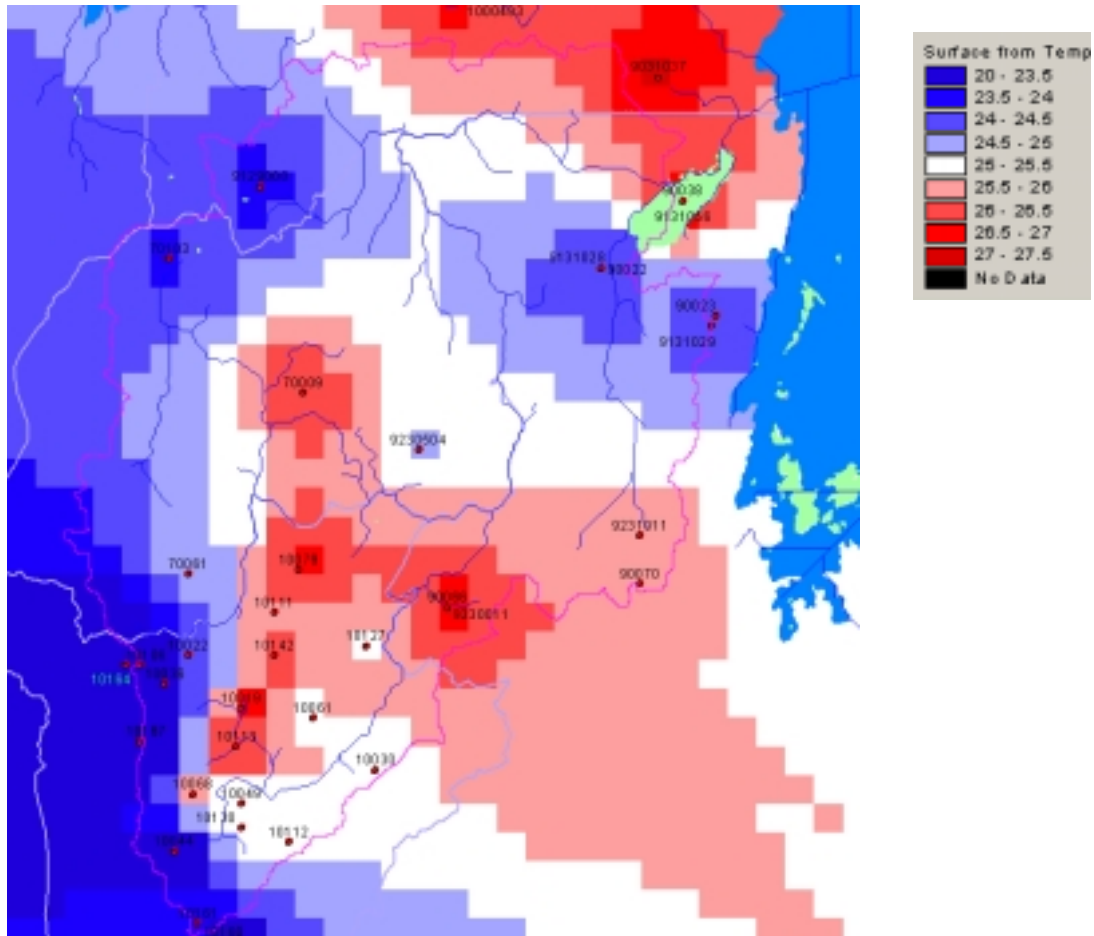


Figure 5.18: Spatial Analysis of Average Annual Temperature

Four stations near Lake Victoria demonstrated much colder AAT values than the surrounding stations. The temperature-elevation function was developed without these stations (Figure 5.19).

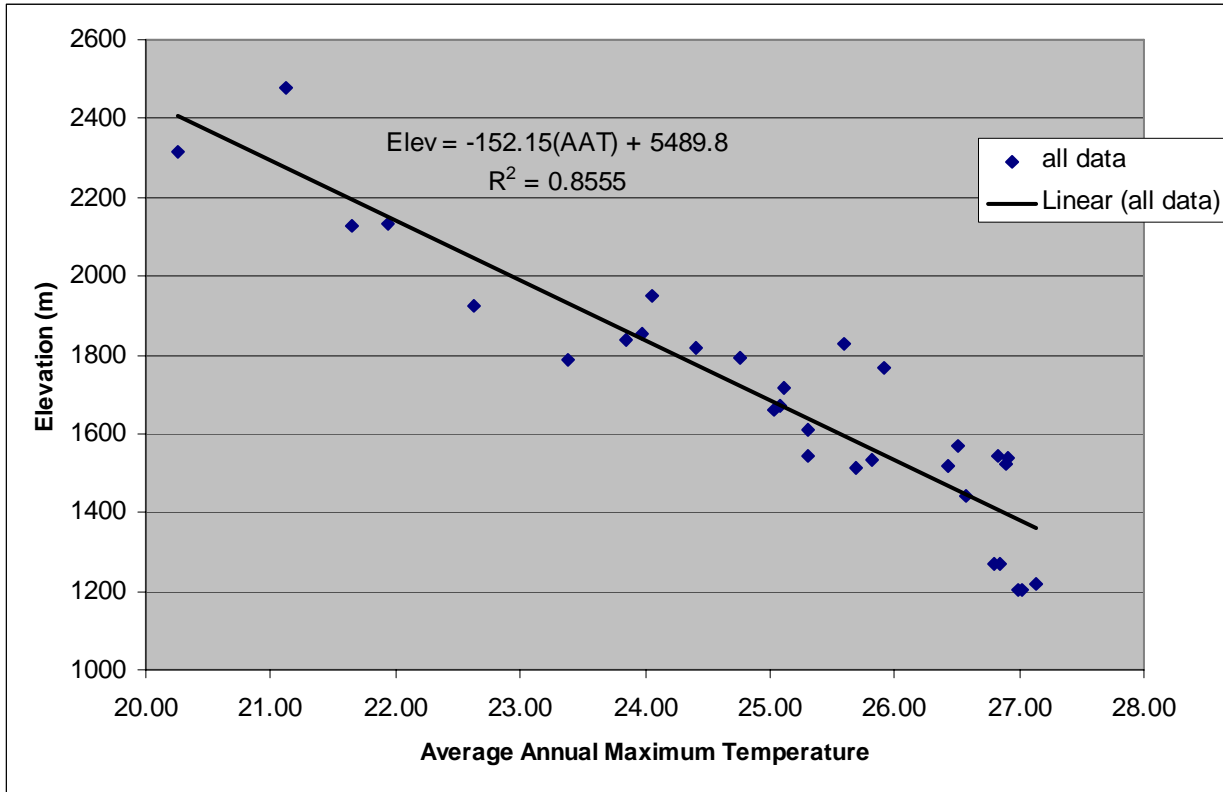


Figure 5.19: Temperature-elevation Function

Similar to the method described in the previous precipitation section, a normal weighted temperature series was developed. The temperature-elevation function was used to develop normal weight values (N) for each grid and for each temperature station. The monthly and 10-day climatologies along with the normal weight values were used in a inverse distance weight algorithm to develop a time series for each grid. These time series were averaged to develop a single temperature time series for the entire basin.

Results

Since climatologies were used the temperature input time series repeats each year. Figure 5.20 and 3.21 show the resulting monthly and 10-day input time series for a single year. This series was repeated from 1960 to 1977 to develop the input file for potential evapotranspiration. The potential evapotranspiration file was used as input for the hydrologic model.

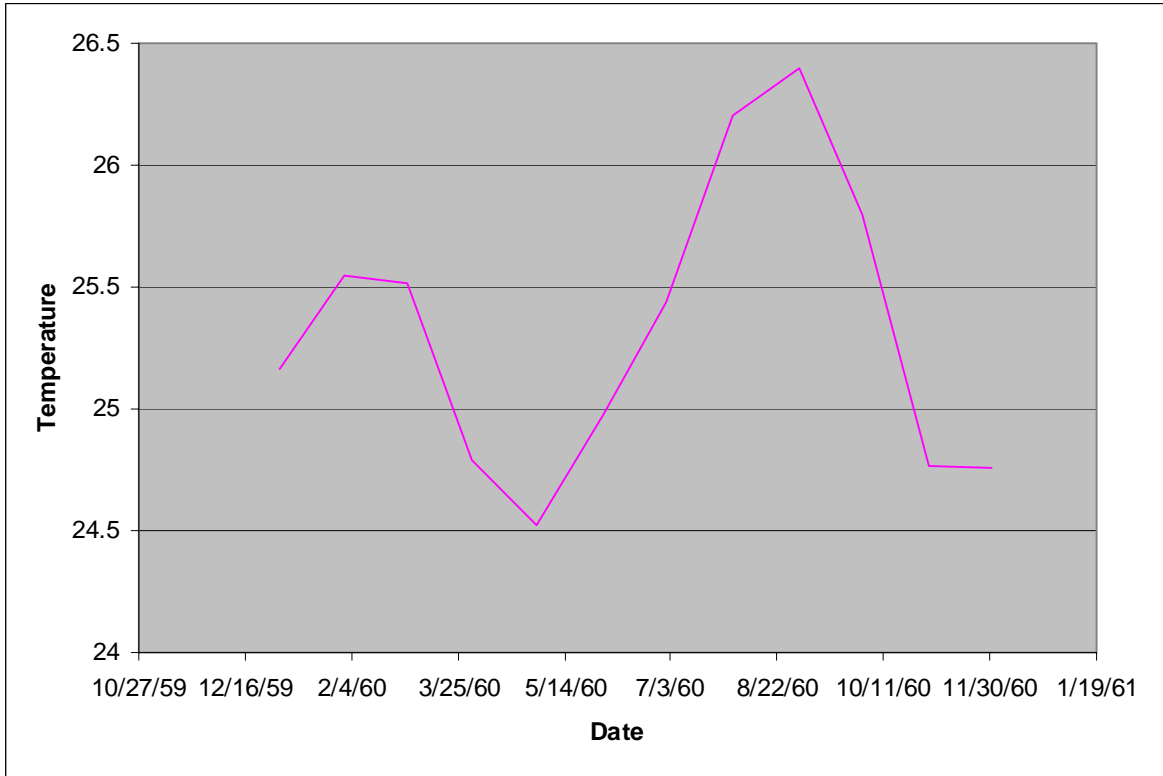


Figure 5.20: Monthly Temperature

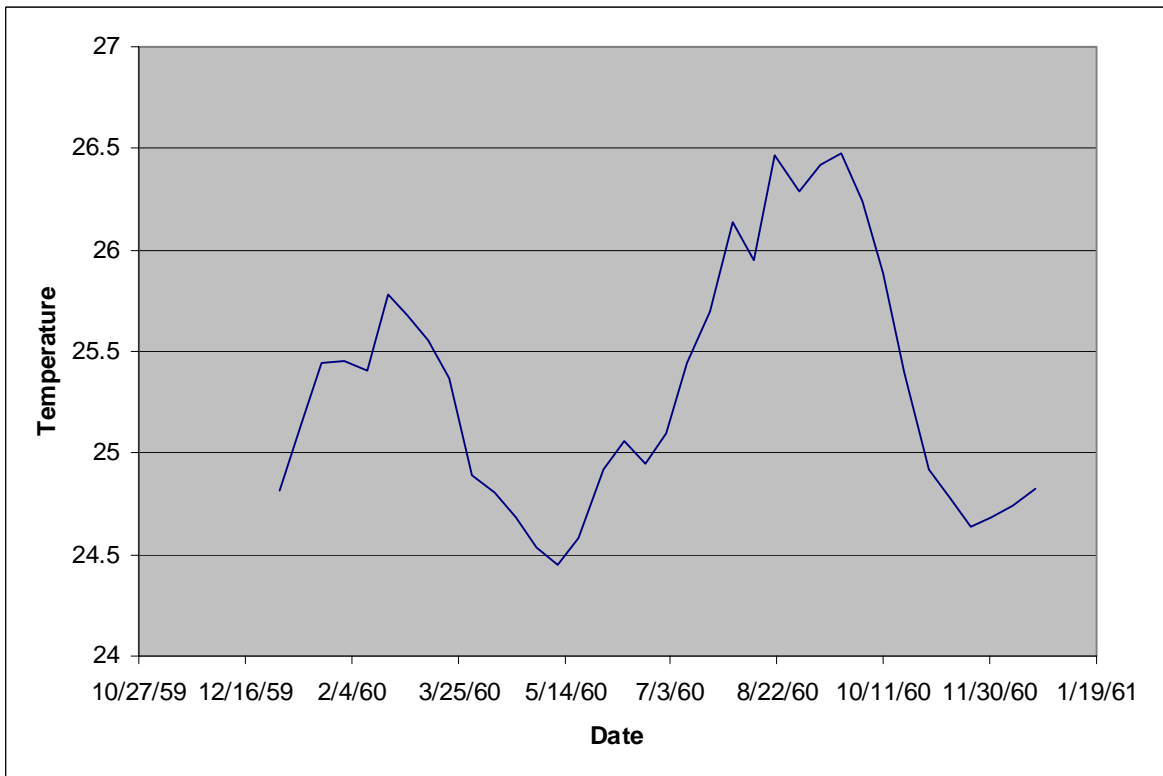


Figure 5.21: 10-day Temperature

5.3 Streamflow Record

Station Selection

Streamflow data was quite limited within the Kagera basin. While there are many gaging stations located within the basin, only two stations were found to have significant data and locations appropriate for modeling the watershed. One station is located in Tanzania near the outlet of the basin to Lake Victoria (Station 90006), while the second station is located in Rwanda (Station 70003) and has an upstream area that drains nearly half of the Kagera watershed.

The upstream Station 70003 contains gage height data between 1956 and 1996. However, the record has many significant issues, including at least 3 years that are duplicated in their entirety, long periods containing a single gage height value, and missing data for months and years. The severity and location of these issues rendered this station a poor choice for the hydrologic model.

The downstream Station 60009 contains streamflow data between 1951 and 1985. Again, this station has significant periods of both repeating and missing data. In this case, however, there is a period between 1960 and 1977 with minimal missing or corrupted data. For this reason, Station 90006 was selected for further data analysis and use in the hydrologic model.

Streamflow Data Analysis and Extension

Initial assessment of Station 90006 indicated a potentially useful data series between 1960 and 1977. Upon closer inspection, some inconsistencies were found in the first year as well as some of the later years. Hydrograph analysis was considered together with the mean areal precipitation data to determine if the flow trends coincided with precipitation trends. As suspected, the trends did not appear to be related for certain time periods. These inconsistencies effectively reduced the usable period to 1961 through 1970; although all years were maintained for data extension so the model could be run with the reduced as well as long periods for further insight into the discrepancies.

Data Extension

Station 90006 required data extension between 1960 and 1977 to fill one- to two-month periods of missing data. There were no other stream gage stations with data in close proximity to this station. However, while Station 70003 is not close to the downstream station, its location does represent the majority of flow that eventually arrives at the Kagera outlet. An assessment of streamflow records for the two stations found that the upstream station contained reasonably good and complete data for the periods the downstream station was missing data.

A multiple linear regression model was developed to relate the upstream flow at Station 70003 with the downstream flow at Station 90006. Station flow data was aggregated into 10-day flows

and modeled using many time steps to account for the significant hydrograph lag time and attenuation as flow progresses downstream. After extensive sensitivity tests on model variables, the final model described downstream flow as a function of the upstream flow at five previous time steps as well as its own flow at one previous time step. The model yielded a correlation coefficient of 0.93.

The regression model was applied to contemporaneous upstream and downstream flow data that was not used during regression. The predicted downstream hydrograph exhibited the major features and trends desired for the purposes of filling and extending one- to two-month data gaps. The model would likely be inappropriate for extending the record for longer periods due to the autoregressive term, which could potentially introduce increasingly more significant errors. Figure 5.22 shows an application of this model for a two-month period of missing data in 1970.

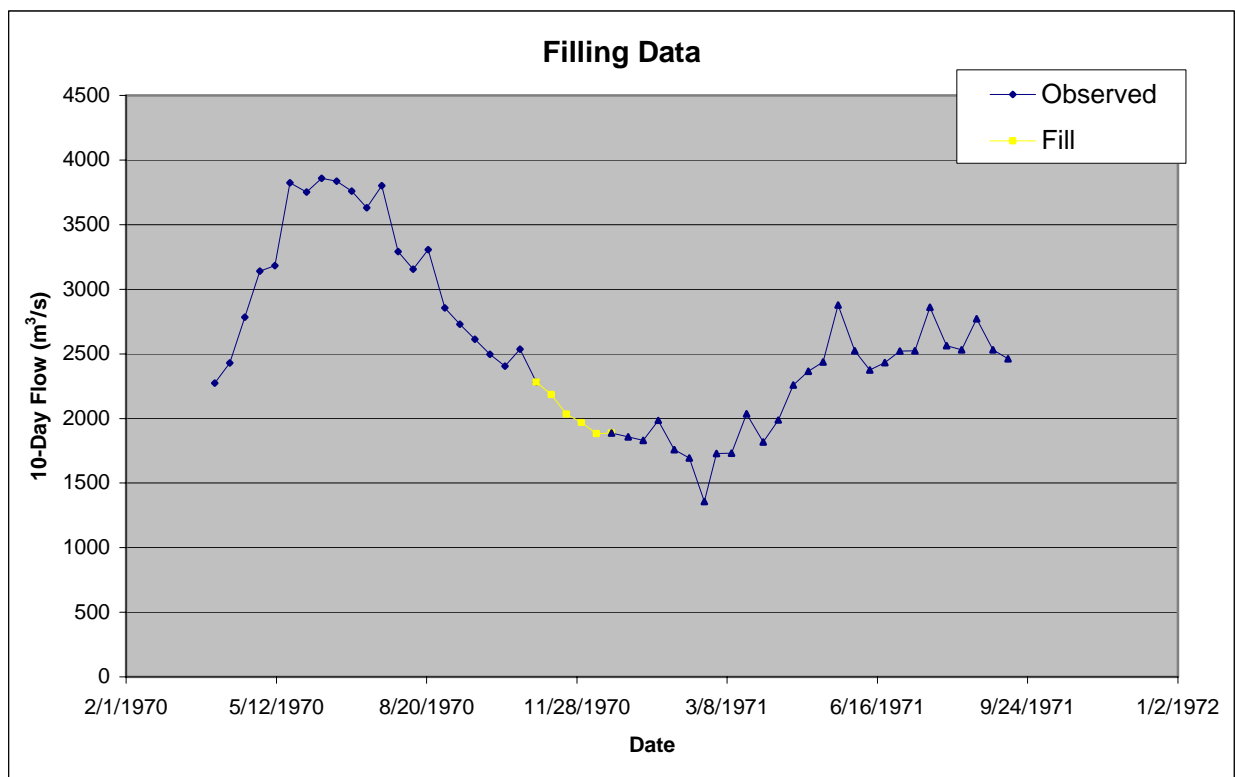


Figure 5.22: Filling Gaps in Streamflow Data

5.4 Model Calibration and Results

The most reliable and continuous period of data (streamflow, precipitation and temperature-based evapotranspiration) extends from 1961 through 1970. Of the 10 years, the first seven years were used to calibrate the hydrologic model, with the remaining three used to verify model results. The model was operated on a 10-day time step. Model results for the verification period are shown in Figure 5.23.



Figure 5.23: Model Verification (1968 – 1970)

The calibrated model was then used to simulate the flow sequence for the entire 10 years (1961 through 1970). Observed and simulated flows are shown and compared in the figure below.

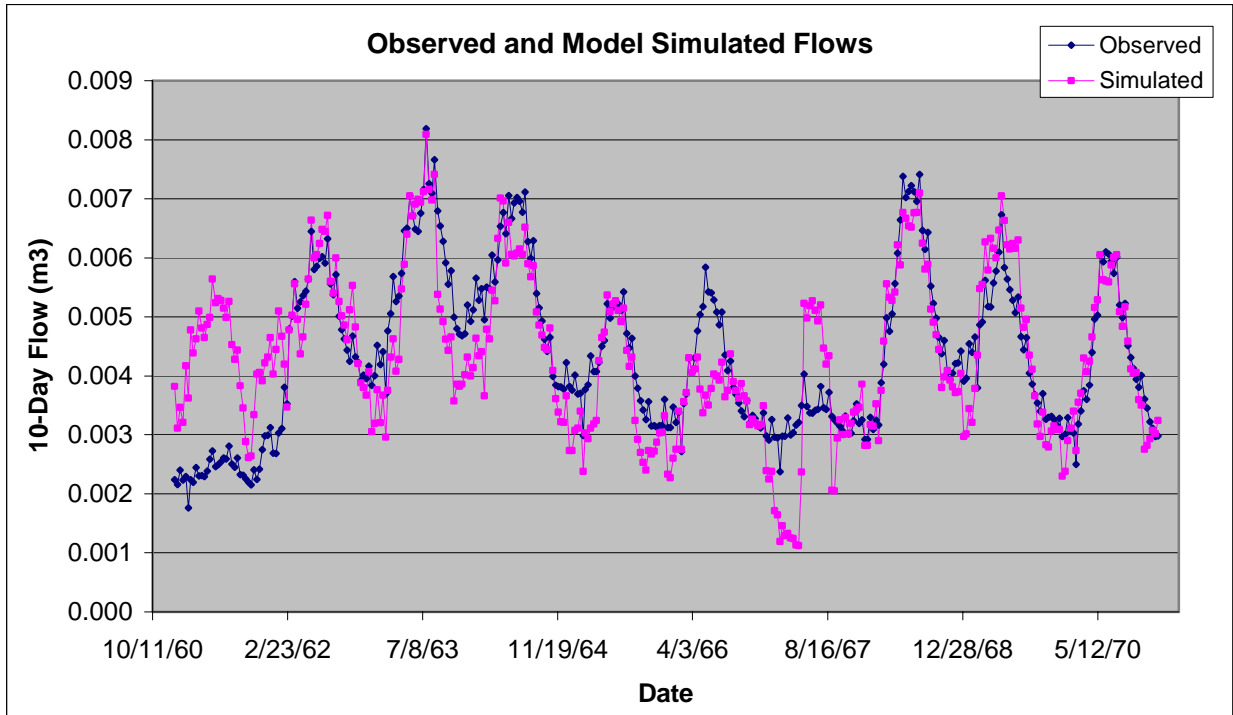


Figure 5.24: Simulated versus Observed Flows (1961 – 1970)

As the figure shows, the model is able to capture several of the peaks very well; yet one mid-range peak and most low flows do not exhibit the actual magnitude of the observed flow. There are many possible reasons for this behavior. As previously mentioned, the data at both streamflow stations considered for this basin was found to have significant periods of corrupted and/or inconsistent data. While the data analysis performed at this stage is extensive, much more investigation and analysis is necessary to identify remaining areas of concern and reconcile discrepancies in the streamflow and precipitation records.

Another issue to consider here is the effect of wetlands on the stream response under high-flow and low-flow conditions. Station 90006 happens to be located within a significant wetland area just upstream of the basin outlet to Lake Victoria. It is likely that the water losses within the wetlands are much higher relative to inflow volume under low-flow conditions than high-flow conditions. With only seven years for calibration and three for verification, there is insufficient data for the model to capture such behavior. Again, further investigation and analysis is required to fully assess the issues within the basin.

The Kagera basin illustrates the contrast in complexity from basin to basin. While the Nzoia basin was calibrated with 10 years of data, just three more than Kagera, the hydrologic behavior over the record is much more regular and consistent with respect to annual and interannual flow and with respect to the precipitation inputs. The Kagera basin, on the other hand, exhibits much more complicated hydrograph behavior. Further, the wetlands near the mouth of the basin add interesting and challenging aspects to the watershed analysis. Country hydrologists will surely have valuable input and contributions in the evaluation and resolution of these issues. Such participation will lead to much improvement and understanding of the unique aspects of each basin.

6. Outstanding Data Issues

While the Nile DST database is extensive, there exist a number of outstanding data issues, which must be addressed before much of the data can be appropriately used. The issues span possible data entry errors, corrupted data, mislabeled data, incomplete datasets, and conflicting trends. This section will highlight a few such issues.

As mentioned previously, in many cases contemporaneous stage-discharge data are available in the database and can be used to develop rating curves. In some instances, graphical analysis of the data indicates that multiple rating curves exist and are embedded within a single set. Figure 6.1 below illustrates one such case. As most of the data points appear to fall directly on a trend line, development of another rating curve from this set may be inappropriate due to the overwhelming influence of points generated from previous equations. Further, there is no accompanying documentation describing the intended use of multiple apparent rating curves. Therefore, it is unclear which curve should be used under which circumstances or whether one curve is more accurate than another. This case presents an excellent opportunity to benefit from the expertise and background knowledge of local engineers.

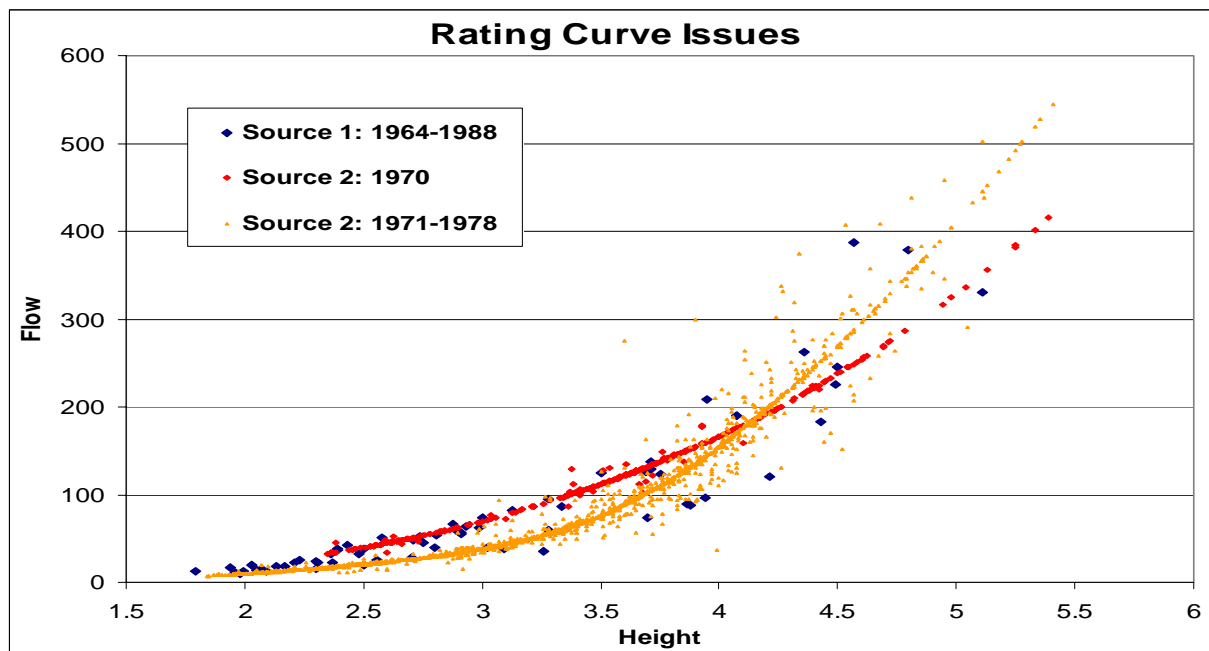


Figure 6.1: Conflicting Stage-Discharge Data

In other cases, where streamflow has been provided explicitly, close hydrograph examination uncovers problems with repeating or unlikely trends. The hydrograph shown below in Figure 6.2 is an example where data is repeated for three different years within a single decade of record. This was found during comparison of two nearby stations for the purpose of evaluating hydrologic consistency of data.

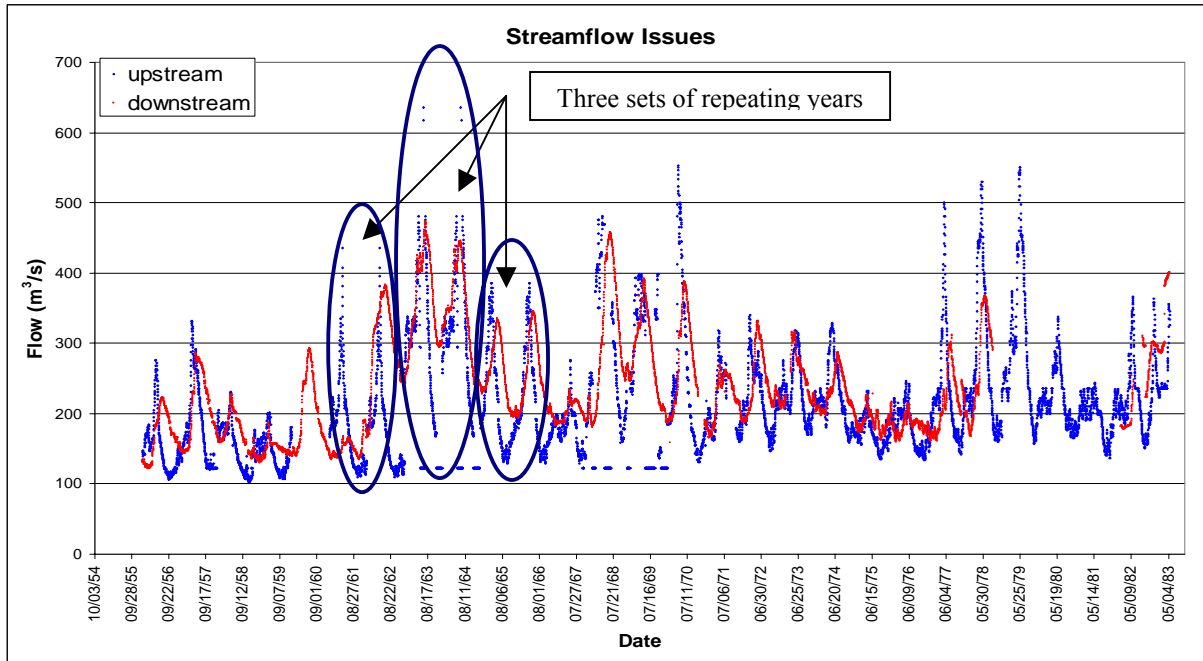


Figure 6.2: Repeating Data (blue series)

The data visualization tool in the Nile DST is used extensively during station identification and selection for model input. Using this tool, graphical analysis of station data sometimes reveals inconsistencies as well as unlikely values within the datasets. Figure 6.3 below is an example of station data for minimum daily temperature. This dataset shows a clear break in trend and raises suspicions that the latter portion of the set may be unreliable.

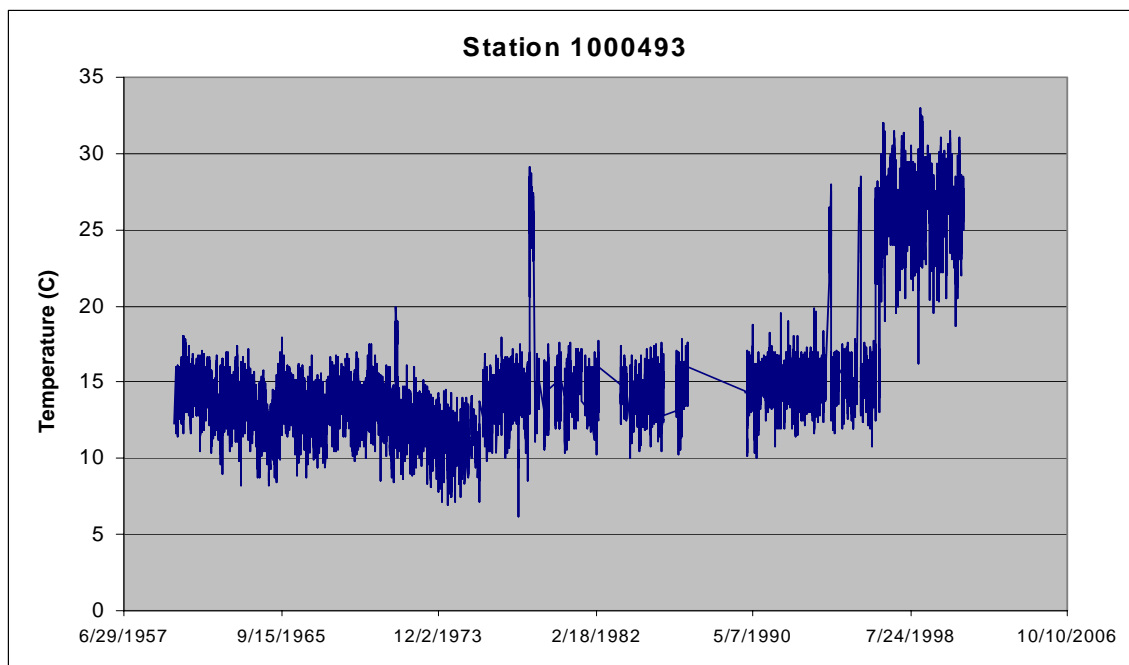


Figure 6.3: Minimum Daily Temperature

Analysis of data for a station near the station shown above shows much higher minimum temperature data (Figure 6.4). In fact, the data in this set are more consistent with maximum temperatures for this area, which suggests that the data may be mislabeled. In the absence of additional information, this station data is unreliable and inappropriate for model input.

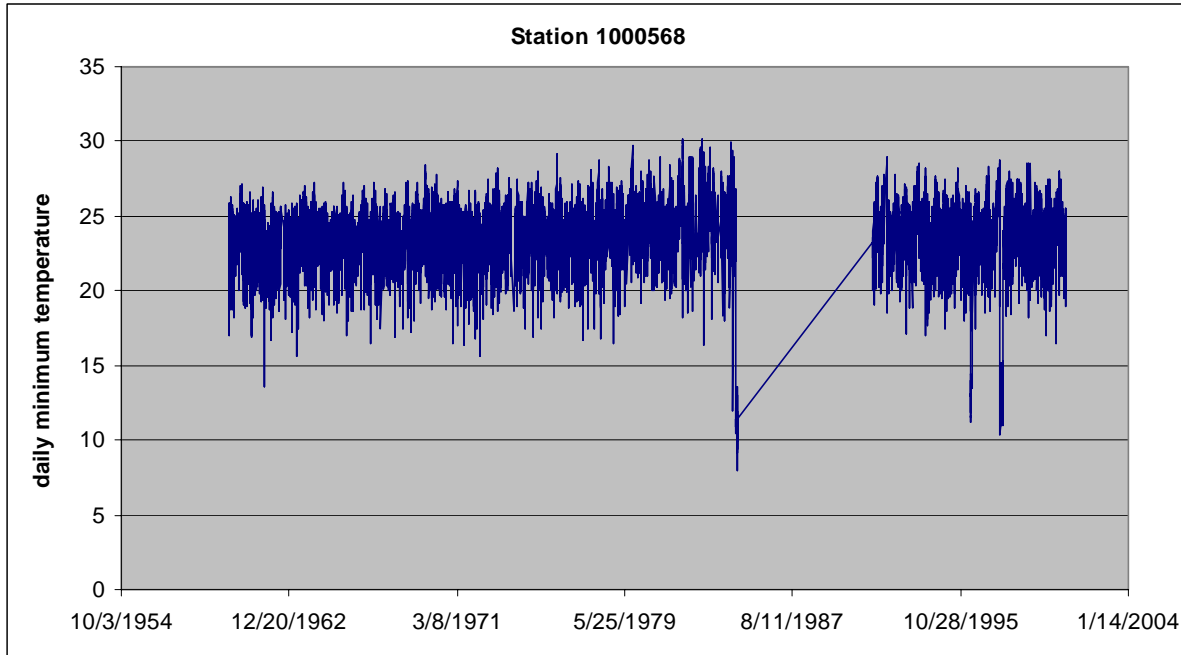


Figure 6.4: Minimum Daily Temperature Inconsistent With Neighboring Stations

Station data is also checked for spatial consistency using a geographic information system (GIS). A region southwest of Lake Victoria is shown below (Figure 6.5) with color-coding based upon average temperature with respect to elevation zones. As can be seen, a station near the center of the image indicates an average temperature that does not coincide with the regional temperature-elevation trend. It is unclear whether this anomaly represents the true climatic behavior near the station or is the result of erroneous data.

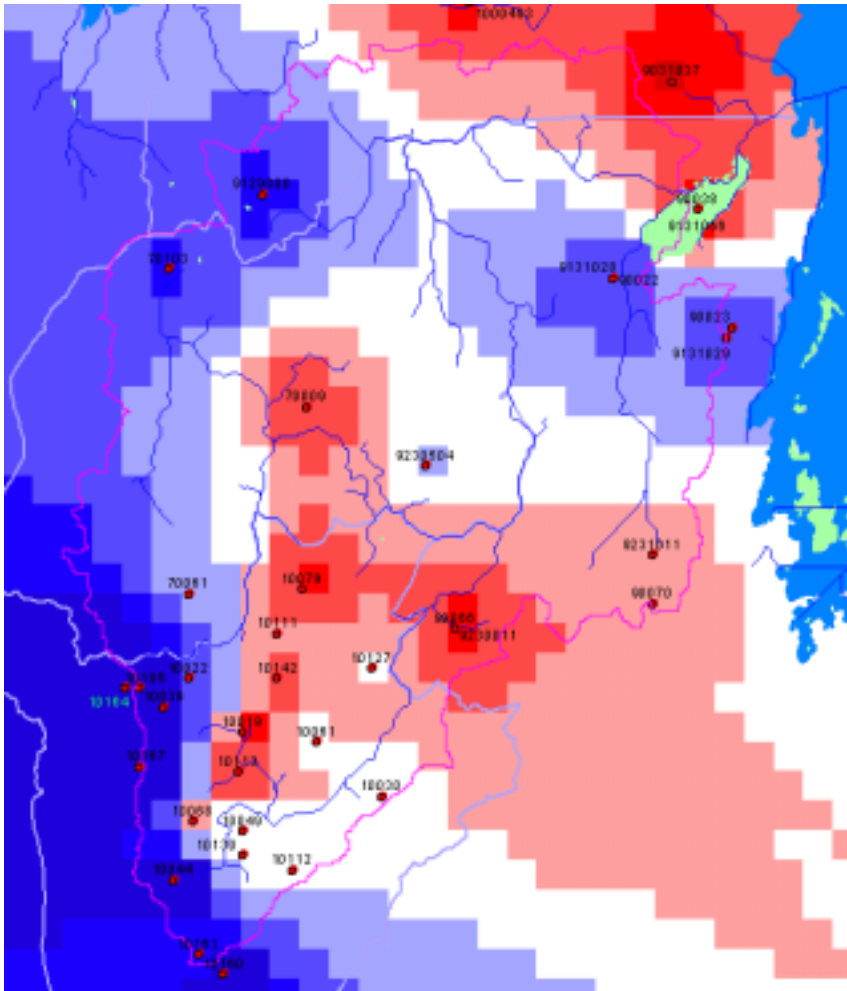


Figure 6.5: GIS Analysis of Spatial Temperature Trends

Spatial evaluation of station distribution is important in determining whether the available data represents an adequate cross section of a particular region. The image below (Figure 6.6) shows GIS representation of precipitation stations west of Lake Victoria. In this case, there is a significant area, near the center of the image, lacking station coverage. The absence of data in such a large region may significantly impair characterization of basin precipitation.



Figure 6.6: Spatial Distribution of Precipitation Stations West of Lake Victoria

In cases where there is adequate station coverage and significant periods of overall record, other issues may still exist. Figure 6.7 shows systematically missing data for one of many stations identified with the same behavior within a region. For these stations there is a very long set of data spanning the 1930's through the 1980's. However, each year from the 1930's through the 1970's is missing data for August through December. The half-year datasets for the precipitation stations in this region effectively render the stations unusable.

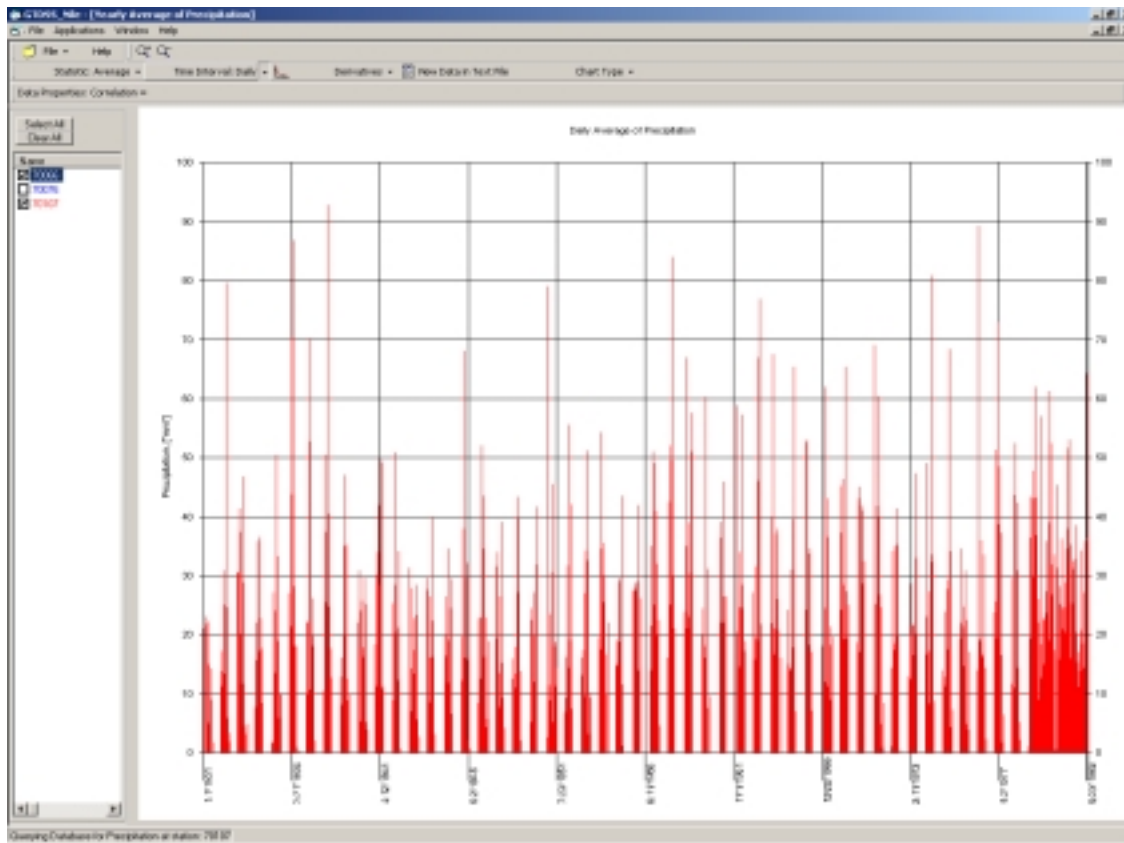


Figure 6.7: Precipitation Station Showing No Data Between August and December

This section has detailed some of the data issues that have been encountered during hydrologic analysis. It is possible that many of these issues may be addressed and resolved through work with country hydrologists and engineers. Such persons may have insight and additional information that allows for proper use, and repair if necessary, of currently available data. Such a cooperative analysis of hydrologic data may greatly improve the quality, as well as quantity, of data available for hydrologic modeling.